

Copyright (c) 1993-1999 Compugen Ltd.

OM protein: protein search, using sw model

Run date: July 21 2002, 17:27:15 : Search time 37.03 seconds
(with alignment)
413 688 Million cell updates/sec

Title: US-09-889-592-2

Perfect score: 1632

Sequences: 1 MPMQSVTRASTG: IVK

Scoring table: BIOSUM 4

Gapop 10.0, Gapext 0.0

Searched: 105224 seqs, 387,750 residues

Total number of hits satisfying criteria: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 characters

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	Id	Description
1	99.5	6.1	501	1	Q4770 drosophila
2	99.5	6.0	502	1	Q4771 drosophila
3	96.5	5.9	681	1	Q17407 caenorhabdi
4	91	5.6	1497	1	Q17407 caenorhabdi
5	87.5	5.4	751	1	Q17407 caenorhabdi
6	87	5.3	346	1	Q17407 caenorhabdi
7	86	5.3	1480	1	Q17407 caenorhabdi
8	85.5	5.2	748	1	Q17407 caenorhabdi
9	84	5.1	660	1	Q17407 caenorhabdi
10	84	5.1	1468	1	Q17407 caenorhabdi
11	83.5	5.1	225	1	Q17407 caenorhabdi
12	83	5.1	325	1	Q17407 caenorhabdi
13	83	5.1	660	1	Q17407 caenorhabdi
14	82.5	5.1	130	1	Q17407 caenorhabdi
15	81.5	5.0	561	1	Q17407 caenorhabdi
16	81.5	5.0	645	1	Q17407 caenorhabdi
17	81	5.0	265	1	Q17407 caenorhabdi
18	80.5	4.9	509	1	Q17407 caenorhabdi
19	80.5	4.9	348	1	Q17407 caenorhabdi
20	80.5	4.9	1916	1	Q17407 caenorhabdi
21	80	4.9	1479	1	Q17407 caenorhabdi
22	80	4.9	1784	1	Q17407 caenorhabdi
23	79.5	4.9	989	1	Q17407 caenorhabdi
24	79.5	4.9	382	1	Q17407 caenorhabdi
25	79	4.8	1524	1	Q17407 caenorhabdi
26	78.5	4.8	1067	1	Q17407 caenorhabdi
27	78.5	4.8	4238	1	Q17407 caenorhabdi
28	78	4.8	391	1	Q17407 caenorhabdi
29	77.5	4.8	132	1	Q17407 caenorhabdi
30	77.5	4.8	1790	1	Q17407 caenorhabdi
31	77.5	4.7	191	1	Q17407 caenorhabdi
32	77.5	4.7	492	1	Q17407 caenorhabdi
33	77.5	4.7	1316	1	Q17407 caenorhabdi
34	77.5	4.7	1316	1	Q17407 caenorhabdi

Result No.	Score	Query Match	Length	Id	Description
35	77	4.7	519	1	Q17407 caenorhabdi
36	77	4.7	821	1	Q17407 caenorhabdi
37	76.5	4.7	410	1	Q17407 caenorhabdi
38	76.5	4.7	685	1	Q17407 caenorhabdi
39	76.5	4.7	826	1	Q17407 caenorhabdi
40	76.5	4.7	1787	1	Q17407 caenorhabdi
41	76.5	4.7	2388	1	Q17407 caenorhabdi
42	76	4.7	412	1	Q17407 caenorhabdi
43	76	4.7	1592	1	Q17407 caenorhabdi
44	75.5	4.6	487	1	Q17407 caenorhabdi
45	75.5	4.6	507	1	Q17407 caenorhabdi

ALIGNMENTS

Result No.	Score	Query Match	Length	Id	Description
35	77	4.7	519	1	Q17407 caenorhabdi
36	77	4.7	821	1	Q17407 caenorhabdi
37	76.5	4.7	410	1	Q17407 caenorhabdi
38	76.5	4.7	685	1	Q17407 caenorhabdi
39	76.5	4.7	826	1	Q17407 caenorhabdi
40	76.5	4.7	1787	1	Q17407 caenorhabdi
41	76.5	4.7	2388	1	Q17407 caenorhabdi
42	76	4.7	412	1	Q17407 caenorhabdi
43	76	4.7	1592	1	Q17407 caenorhabdi
44	75.5	4.6	487	1	Q17407 caenorhabdi
45	75.5	4.6	507	1	Q17407 caenorhabdi

Result No.	Score	Query Match	Length	Id	Description
35	77	4.7	519	1	Q17407 caenorhabdi
36	77	4.7	821	1	Q17407 caenorhabdi
37	76.5	4.7	410	1	Q17407 caenorhabdi
38	76.5	4.7	685	1	Q17407 caenorhabdi
39	76.5	4.7	826	1	Q17407 caenorhabdi
40	76.5	4.7	1787	1	Q17407 caenorhabdi
41	76.5	4.7	2388	1	Q17407 caenorhabdi
42	76	4.7	412	1	Q17407 caenorhabdi
43	76	4.7	1592	1	Q17407 caenorhabdi
44	75.5	4.6	487	1	Q17407 caenorhabdi
45	75.5	4.6	507	1	Q17407 caenorhabdi

CC -1- FUNCTION: MAY BE INVOLVED IN THE METABOLISM OF INSECT HORMONES AND
 CC IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) -> ROH +
 CC oxidized flavoprotein + H(2O).
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
 CC (Potential).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb.sib.ac.at/submit/>
 CC or send an email to license@isb.sib.ch)
 CC -----
 DR EMBL: AF003813; AAF58190.1; --
 DR FlyBase: FB00015714; CytoP417.
 DR InterPro: IPR001128; CYP_P450.
 DR Pfam: PF00067; p450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 DR oxidoreductase; Monooxygenase; Membrane; Hemol; Microsome;
 KW Endoplasmic reticulum; Hypothetical protein
 FT BINDING: 444 444 HEME (BY SIMILARITY).
 SQ SEQUENCE: 501 AA; 58208 MW; 00E0E8364AE40D41 CR664;

Query Match 6.0%; Score 99.5; DB 1; Length 501;
 Best Local Similarity 24.1%; Pred. No. 0.21;
 Matches 70; Conservative 42; Mismatches 90; Indels 99; Gaps 18;
 QY 27 PHAKVQAKQAQIDPEEE-----LSVKPKMKVNIHINIDPQQAQ-----YR 69
 DB 200 PKAEFVSQKRAITLRYKYNMIDETFGCPKTSRRRIKINCEAFHFKIVKRETDYR 259
 QY 70 LLENEOTEFISDPSCLR-----ISDKYLIAWVLAAYPKRAAGLYTSEYTHN 116
 DB 260 LRTEKNDP--MUSLEMTKNSQNSHEDGTFNELLQAQIFF-VAGFEISS-IMG 314
 QY 117 FFVALY-LANDMDEDEEYKYEIPWALGDSWRELPQFLRIEDP---WAKMN---YR 168
 DB 315 F--ALVELARN--QDVQD-----KLREFGNVEKIKNEFTYE 348
 QY 169 AVVSRCCDVMKDPTHAWLKDIP-MHUSCAMCYLPNEHFFPPEPPEPPEPPEPPEPPEP 227
 DB 349 GTRKMKYLEQVVMET-----LKKYVLAHGRM TDTDFSPEDPKYFIARGLIV- 396
 QY 228 KAGVCDGSGWSHNNSSPQEIF---HYTNREWSQELMLPTELLDPECT 275
 DB 397 ---VTPALGTHYDDIYDPEIKPKPEPTDE-----ELAAISCT 433

RESULT 2
 CHAN_DROME STANDARD; PRT: 502 AA.
 AC QV771;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable cytochrome P450 6a23 (EC 1.14.1.1) (CYP6A23).
 GN CYP6A23 OR CG10242.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RC MEDLINE=20196006; PubMed=10741132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Garayue J.D.,
 RA Abamailides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

George R.A., Lewis S.E., Richards S., Ashburner M., Bruchman S.N.,
 SA Sutton G.S., Wortman J.K., Yandell M.D., Zhang Q., Chen L.X.,
 SA Anderson R.S., Rogers Y., Hultine M., Chapple M., Pfeiffer B.L.,
 SA Wan K.H., Doyle C., Baxter E.G., Holt C., Nelson C.R., Miklos G.L.,
 SA Abril J.F., Adayani A., An H., J., Andrews P., Binkov B., Bohn P.,
 SA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 SA Beeson K.Y., Benos P.V., Beran J., Bhattacharya D., Bickel S.,
 SA Borokova D., Botchan M.R., Boucek J., Brockstein P., Butler P.,
 SA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra P.,
 SA Cherry J.M., Cawley S., Bahke C., Davenport L.R., Davies P.,
 SA de Pablos R., Delecher A., Deng Z., Mays A.D., New T., Dierckx M.,
 SA Dodson K., Dong L.E., Dones M., Duran Rocha S., Dunkov B., Dunn P.,
 SA Durbin K.J., Evangelista G.C., Ferraz C., Ferreira S., Fleischmann W.,
 SA Essler R., Gabriellian A.E., Gary R.S., Gehlert W.M., Glasser A.,
 SA Glodok A., Gong F., Gottlieb J.H., Gu Z., Guan L., Harris M.,
 SA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Henik J.,
 SA Hostin D., Houston K.A., Howland T.J., Wei M., Huettemann C.,
 SA Jalili M., Falush F., Karpe P.H., Kozlov A., Kozlov A., Kozlov A.,
 SA Kimmel P.E., Kodira G.D., Kraft G., Kravitz S., Kulp D., Lai Z.,
 SA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 SA Liu X., Matten R., McIntosh L.C., McLeod M.P., McPherson D.,
 SA Morkov G., Milshina N.V., Mobarry C., Morris J., Moshell A.,
 SA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 SA Nelson D.P., Nelson K.A., Nixon K., Nusskern H.K., Pauley J.M.,
 SA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 SA Reinert K., Remington K., Saunders R.D., Scheibel P., Shen B.,
 SA Shue H.C., Siden Kamos L., Simpson M., Skupski M.P., Smith D.,
 SA Spiet E., Spradling A.C., Stapleton M., Strong K., Sun E.,
 SA Swirskas K., Tetter R., Turner R., Venter L., Wang A.H., Wang X.,
 SA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 SA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 SA Ye J., Yen R.F., Zayori J.S., Zhan M., Zhang L., Zhao X., Zhou L.,
 SA Zhou X.H., Zhou J.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.,
 SA Gibbs R.A., Myers E.W., Rubin G.M., Venter L., Smith H.,
 SA *The genome sequence of Drosophila melanogaster.*
 SC Science 287:2165-2195(2000).
 EN [2]
 CC CONCEPTUAL TRANSLATION
 RA Neilson R.
 CC Unpublished observations (SEP 2000).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE METABOLISM OF INSECT HORMONES AND
 CC IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) -> ROH +
 CC oxidized flavoprotein + H(2O).
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
 CC (Potential).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -1- CAUTION: REF-1 SEQUENCE DIFFERS AT SEVERAL POSITIONS FROM HAI
 CC SHOWN DUE TO DIFFERENCES IN THE PREDICTION OF SPICE SITES.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb.sib.ac.at/submit/>
 CC or send an email to license@isb.sib.ch)
 CC -----
 DR EMBL: AF003813; AAF58190.1; ALL SEQ.
 DR FlyBase: FB00015714; CytoP417.
 DR InterPro: IPR001128; CYP_P450.
 DR Pfam: PF00067; p450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 DR oxidoreductase; Monooxygenase; Membrane; Hemol; Microsome;
 KW Endoplasmic reticulum; Hypothetical protein
 FT BINDING: 444 444 HEME (BY SIMILARITY).
 SQ SEQUENCE: 502 AA; 58089 MW; 14A1676AM298496 CR664;

Query Match 6.0%; Score 99.5; DB 1; Length 502;
 Best Local Similarity 21.9%; Pred. No. 0.21;
 Matches 75; Conservative 34; Mismatches 90; Indels 130; Gaps 18;

FT	SITE	514	516	DEAD	POX
SQ	SEQUENCE	1466	AA: 146240	MM: 78860740	FF: 0065, ch6: 64
	Quota Match		5, 19;	Score: 84;	DB: 1;
	Best Local Similarity		20, 68;	Prod. No: 18;	
	Matches	74;	Conservative	60;	Mismatch: 12;
					Index: 96;
					Gap: 1
QY	14	KAVINCHODHA	VVYAPPA	QAPQ	DEPE
		1111	1111	1111	1111
		1111	1111	1111	1111
DB	1072	KUCHEKKAHA	AVAKAKA	KSA	KRANAKA
		1111	1111	1111	1111
		1111	1111	1111	1111
QY	78	FELISACST	PISD	KYLAMALAV	KRAACH
		1111	1111	1111	1111
		1111	1111	1111	1111
DB	1126	KULAKASIS	SRVETAKYUN	INSINAKUP	KESANNIDE
		1111	1111	1111	1111
		1111	1111	1111	1111
QY	136	YEYETWAL	ADISWRELE	TO	FIERDOO
		1111	1111	1111	1111
		1111	1111	1111	1111
DB	1185	IGVRYGYS	NWTO	KOPEP	PIETIT
		1111	1111	1111	1111
		1111	1111	1111	1111
QY	174	---	---	KCEDEWYS	---
		1111	1111	1111	1111
		1111	1111	1111	1111
DB	1244	CKKVYTA	ILIGKRVY	WAST	YKGLINIK
		1111	1111	1111	1111
		1111	1111	1111	1111
QY	204	YLRREB	---	---	---
		1111	1111	1111	1111
		1111	1111	1111	1111
DB	1304	FELISSHAN	DESKAMAI	DECTAAL	ENNI
		1111	1111	1111	1111
		1111	1111	1111	1111
QY	247	QELPHYIN	RESQEL	MLPDE	LLDP
		1111	1111	1111	1111
		1111	1111	1111	1111
DB	1462	SISAB	---	---	---
		1111	1111	1111	1111
		1111	1111	1111	1111

RESULT	11
CRP_RAB11	
ID	CRP_RAB11
AC	STANDARD
DC	PWT: 22% AA.
DT	P02742;
D1	21-JUL-1986 (Ref. 01, Created)
D2	01-JAN-1988 (Ref. 06, Last sequence update)
D3	16-OCT-2001 (Ref. 40, Last annotation update)
DE	Corrective protein precursor.
GN	PTX1 OR CRP.
GS	Cryptosporidium parvum (Rabbit).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Eumetazoa; Tetrapoda; Artiodactyla;
OX	NCBI_taxid=9486;
RI	11
RP	SEQUENCE FROM N.A.
RC	STRAIN-NEW ZEALAND WHITE; TISSUE-Liver;
RF	MEDLINE=87101076; PubMed=4026463;
RA	Wu S., J. Biol. Chem. 261, 5473-5479(1986).
RE	"Cloning and characterization of the gene for rabbit C reactive protein."
RT	Biochemistry 25:7834-7839(1986).
RI	12
RP	SEQUENCE FROM N.A.
RF	MEDLINE=86168292; PubMed=3007566;
RA	Sytkin C., Gottschlich E., Liu L. Y.,
RE	"Rabbit C reactive protein: biosynthesis and characterization of cDNA clones."
RT	J. Biol. Chem. 261:5473-5479(1986).
RI	13
RP	SEQUENCE OF 21-225.
RF	MEDLINE=84056861; PubMed=6794715;
RA	Wang C. M., Nguyen N. Y., Yoshida K., Kobayashi T., Yip
RE	"Primary structure of rabbit C reactive protein."
RT	J. Biol. Chem. 257:13610-13615(1982)
CC	1-1 FUNCTION: CRP DISPLAYS SEVERAL FUNCTIONS ASSOCIATED WITH DEFENSE: 1) PROMOTES AGGLUTINATION, BACTERIAL CAPSULE SWELLING,
CC	PHAGOCYTOSIS, & COMPLEMENT FIXATION THROUGH ITS CAUCTION DOMINANT
CC	BINDING TO PHOSPHORYLCHOLINE.
CC	1-1 SUBUNIT: HOMOPENTAMER. PENTAXIN (OR PENTAXIN) HAVE A DISCREET
CC	ARGUMENT OF 5 NONCOVALENTLY BOUND SUBUNITS.

CC -1 INDUCTION: THE CONCENTRATION OF THE IN PLASMA INCREASES GREATLY
 CC DURING ACUTE PHASE RESPONSE TO TISSUE INJURY OR INFLAMMATION.
 CC -1 MISCELLANEOUS: ASP-6 (AR-76, ARC-7) AND JLU-81 MAY BE INVOLVED
 CC IN THE CALCIUM-DEPENDENT BINDING OF PHOSPHORYLCHOLINE, A PROPERTY
 CC THAT MAY BE IMPORTANT FOR THE BIOLOGICAL FUNCTION OF THIS PROTEIN.
 CC -1 SIMILARITY: BELONGS TO THE PENTAXIN FAMILY.
 CC THIS SWISS-PROT entry is updated: 1) produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announcements>
 CC or send an email to license@isb.ch).
 DR EMBL: M14538; AAA75403.1; 225 AA; 2549 MW; 60.5 kDa; 14031428; CRCL1.
 DR EMBL: L47237; AAA75404.1; 225 AA; 2549 MW; 60.5 kDa; 14031428; CRCL1.
 DR EMBL: M1497; AAA3106.1; 225 AA; 2549 MW; 60.5 kDa; 14031428; CRCL1.
 DR PIR: A01202; CJRR.
 DR PIR: A2563; A2560.
 DR PIR: A2563; A2560.
 DR HSS: P02741; ICBV.
 DR InterPro: IPR001759; Pentaxin.
 DR Pfam: PF00354; Pentaxin_1.
 DR PRINTS: PF00895; PENTAXIN.
 DR ProDom: PD002153; Pentaxin_1.
 DR SMART: SM00159; PTX_1.
 DR ProSITE: PS00289; PENTAXIN; 1.
 KW Acute phase; Plasma; Calcium; Pentaxin; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 225
 FT DOMAIN 21 225
 FT DISULFID 55 116
 FT VARIANT 62 62
 FT VARIANT 89 89
 FT CONFLICT 10 10 (IN REF. 2)
 FT CONFLICT 46 46 (IN REF. 3)
 FT CONFLICT 50 50 (IN REF. 3)
 FT CONFLICT 76 76 (IN REF. 3)
 FT CONFLICT 84 88 (IN REF. 3)
 FT CONFLICT 91 91 (IN REF. 3)
 FT CONFLICT 93 101 (IN REF. 3)
 FT CONFLICT 104 108 (IN REF. 2)
 FT CONFLICT 167 167 (IN REF. 3)
 FT CONFLICT 177 177 (IN REF. 3)
 FT CONFLICT 193 193 (IN REF. 3)
 SQ SEQUENCE 225 AA; 2549 MW; 60.5 kDa; 14031428; CRCL1.

Query Match 5.1%, Score 83, DB 1, Length 225,
 Host Local Similarity 23.8%, Prod. No. 2 27, Indels 79, Gaps 15,
 Matches 64; Conservative 30; Mismatches 67.
 QY 79 ELMSLSCLRSKYLIAMVLAVERKAASHV SE - YITMN -----FFVALYLAN 125
 DB 8 ELVLSFNNMSQ-----ACMHKATVUKESONSVSLSNQLKPKAKFTVCLYPT 60
 QY 129 MEHREHREYHLEWAGTWFEP LPTETTKAK MVRKAVVTPDCLVMKX 142
 DB 61 ELMSLTP-----GYSF-YATP-----PQVNFTH-----LWKLGYNSVS-----DETTRKVS 106
 QY 184 GPTH-----W-----AWLFRPMHSGAMPGYL-----RNEDDP-----EP 213
 DB 167 ELVYDTHICASWSESTIAHWVGRIMAKS EYVLEIASLIDQWQSPQSEF 166
 QY 214 ELHLEFASVILHAKAVHSGAVVGRVYRQV LPHYHINSEVLELL 242
 DB 167 ELVLSL-----VPLGNVAVVCAVSPRENTIYATGSENVL 264

RESULT 32
 CC10_ANAL

ID CC10_ANAL STANDARD; PRI: 497 AA;

A F49827;
 Q 1 FEB 1995 (Rel. 31, Created)
 Q 61 FEB 1995 (Rel. 31, Last sequence update)
 Q 16-OCT-2001 (Rel. 40, Last annotation update)
 CC Cell division control protein 10.
 CC CDCL10.
 CC Candida albicans (Yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 CC NR01_taxid=5476;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC SYPAIN-0742;
 CC MEDLINE:94201183; PubMed:8152419;
 CC Didomenico B.J., Brown N.H., Lupisella J., Greene J.R., Yanko M.,
 CC Koltin Y.,
 CC "Homologs of the yeast neck filament associated genes: isolation and
 CC sequence analysis of Candida albicans CDC3 and CDC10.";
 CC Mol. Gen. Genet. 242:689-694(1994).
 CC -1- FUNCTION: PLAYS A ROLE IN THE CELL CYCLE. INVOLVED IN THE
 CC FORMATION OF THE RING OF FILAMENTS IN THE NECK REGION AT
 CC THE MOTHER-HUD JUNCTION DURING MITOSIS.
 CC -1- SUBCELLULAR LOCATION: PRESENT AT THE HUD NECK DURING CELL
 CC DIVISION (BY SIMILARITY).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT HIGHER LEVELS IN CELLS GROWING
 CC AS HYPHAE THAN IN THOSE GROWING AS BUDDING YEASTS.
 CC -1- SIMILARITY: BELONGS TO THE SEPTIN FAMILY.
 CC THIS SWISS-PROT entry is copyright: it is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announcements>
 CC or send an email to license@isb.ch).
 DR EMBL: Z25870; CAA81090.1; 74
 DR InterPro: IPR000038; GTP_Cell_Div.
 DR Pfam: PF00735; GTP_CDC; 1.
 DR ProDom: PD002565; GTP_Cell_Div; 1.
 KW Cell division; Cell cycle; GTP-binding.
 FT NP_BIND 44 51 GTP (POTENTIAL).
 FT SEQUENCE 357 AA; 40654 MW; 7025008336274942 CRCL4;
 Query Match 5.1%, Score 83, DB 1, Length 457;
 Host Local Similarity 21.7%, Prod. No. 4.2;
 Matches 38; Conservative 35; Mismatches 70; Indels 32; Gaps 7;
 QY 27 PHAVVVARPAQIPEPEELSVFKMVENTIILN-----LQDEEQAYRLENE 74
 DB 117 PLKRYKREQSQYLRKELIACHOKFLADIVKGLIFIPNGGKFLQIDVQALKRLSIA 176
 QY 75 QIQEFLSMDSCLRLSKYLIAMVLAVERKAAGLYSTVITMNFVALYLANDMEEDF--- 141
 DB 177 NVVPLIAKSDSLTDER-----SEPKK--LLSEFMKNINHYDELYEERQL 226
 QY 172 EKAKYLIIFWALGLSWP LITFQLELELIFWAKMKSFAVVRKGLVNSGL 183
 DB 16 OCT 2001 (Rel. 40, Last sequence update)
 QY 16 OCT 2001 (Rel. 40, Last sequence update)
 QY 16 OCT 2001 (Rel. 40, Last annotation update)
 CC Nubulin.
 CC NFB.
 CC Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95257391; PubMed=7739042;
 RA Labelt S., Kolmerer B.;
 RT "The complete primary structure of human nebulin and its correlation
 to muscle structure";
 RL J. Mol. Biol. 248:408-415(1995).
 RN [2]
 RP PARTIAL PRELIMINARY SEQUENCE.
 RX MEDLINE=88284704; PubMed=4397062;
 RA Zeviani M., Darras H.T., Rizzuto K., Salvati G., Botto R.,
 RA Bonilla F., Miranda A.F., Di J., Samitt C., Dickson G., Walsh F.S.,
 RA Dimauro S., Francke U., Schon E.A.;
 RT "Cloning and expression of human nebulin cDNAs and assignment of the
 gene to chromosome 2q31-q42";
 RL Genomics 2:249-256(1988).
 RN [3]
 RP STRUCTURE BY NMR OF 6610-6669.
 RX MEDLINE=98179559; PubMed=9514727;
 RA Politou A.S., Millevol S., Gautel M., Kolmerer B., Pastore A.;
 RT "SH3 in muscles: solution structure of the SH3 domain from nebulin";
 RL J. Mol. Biol. 276:189-202(1998).
 CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MAINTAINING
 THE STRUCTURAL INTEGRITY OF SARCOMERES AND THE MEMBRANE SYSTEM
 ASSOCIATED WITH THE MYOFIBRILS. BIND AND STABILIZE F-ACTIN.
 CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC. LOCATED IN THE THIN FILAMENT
 OF STRIATED MUSCLE.
 CC -1- DISEASE: Defects in NBR are a cause of the autosomal recessive
 form of nemaline myopathy (NEM2).
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN
 CC -1- SIMILARITY: CONTAINS 178 NEBULIN REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.org/amsn/)
 or send an email to license@sib-sb.ch)
 DR EMBL: X84957; CAA58788.1;
 DR EMBL: M19668; AAA59916.1; AL1_SHQ.
 DR EMBL: M19669; AAA59917.1; AL1_SEQ.
 DR PIR: A29979; A29979.
 DR PIR: B29979; B29979.
 DR IUB: IARK; 28-JAN-98.
 DR IUB: INER; 24-DEC 97.
 DR MIM: 161650;
 DR MIM: 256030;
 DR InterPro: IPR000900; Nebulin.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00880; Nebulin_repeat; 143.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00510; NEBULIN.
 DR SMART: SM00227; NBR; 178.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50002; SH3; 1.
 KW Actin-binding; Muscle protein; Cytoskeleton; Repeat; 3d-structure.
 FI REPEAT 76 107 NEBULIN 1.
 FI REPEAT 113 143 NEBULIN 2.
 FI REPEAT 148 178 NEBULIN 3.
 FI REPEAT 183 213 NEBULIN 4.
 FI REPEAT 218 248 NEBULIN 5.
 FI REPEAT 253 283 NEBULIN 6.
 FI REPEAT 289 318 NEBULIN 7.
 FI REPEAT 324 354 NEBULIN 8.
 FI REPEAT 363 393 NEBULIN 9.
 FI REPEAT 398 428 NEBULIN 10.
 FI REPEAT 434 464 NEBULIN 11.
 FI REPEAT 502 542 NEBULIN 12.
 FI REPEAT 537 567 NEBULIN 13.
 FI REPEAT 603 641 NEBULIN 14.
 FI REPEAT 681 711 NEBULIN 15.
 FI REPEAT 749 779 NEBULIN 16.
 FI REPEAT 820 850 NEBULIN 18.
 FI REPEAT 888 923 NEBULIN 20.
 FI REPEAT 924 954 NEBULIN 21.
 FI REPEAT 959 990 NEBULIN 22.
 FI REPEAT 1023 1058 NEBULIN 24.
 FI REPEAT 1028 1058 NEBULIN 25.
 FI REPEAT 1064 1094 NEBULIN 26.
 FI REPEAT 1102 1132 NEBULIN 27.
 FI REPEAT 1137 1167 NEBULIN 28.
 FI REPEAT 1168 1198 NEBULIN 29.
 FI REPEAT 1204 1234 NEBULIN 30.
 FI REPEAT 1237 1267 NEBULIN 31.
 FI REPEAT 1272 1302 NEBULIN 32.
 FI REPEAT 1308 1338 NEBULIN 33.
 FI REPEAT 1346 1376 NEBULIN 34.
 FI REPEAT 1411 1441 NEBULIN 35.
 FI REPEAT 1412 1442 NEBULIN 36.
 FI REPEAT 1448 1478 NEBULIN 37.
 FI REPEAT 1511 1546 NEBULIN 38.
 FI REPEAT 1516 1546 NEBULIN 39.
 FI REPEAT 1552 1582 NEBULIN 40.
 FI REPEAT 1590 1620 NEBULIN 41.
 FI REPEAT 1625 1655 NEBULIN 42.
 FI REPEAT 1656 1686 NEBULIN 43.
 FI REPEAT 1722 1752 NEBULIN 44.
 FI REPEAT 1725 1755 NEBULIN 45.
 FI REPEAT 1760 1790 NEBULIN 46.
 FI REPEAT 1796 1826 NEBULIN 47.
 FI REPEAT 1834 1864 NEBULIN 48.
 FI REPEAT 1869 1899 NEBULIN 49.
 FI REPEAT 1900 1930 NEBULIN 50.
 FI REPEAT 1936 1966 NEBULIN 51.
 FI REPEAT 1969 1999 NEBULIN 52.
 FI REPEAT 2004 2034 NEBULIN 53.
 FI REPEAT 2040 2070 NEBULIN 54.
 FI REPEAT 2078 2108 NEBULIN 55.
 FI REPEAT 2114 2143 NEBULIN 56.
 FI REPEAT 2144 2174 NEBULIN 57.
 FI REPEAT 2180 2210 NEBULIN 58.
 FI REPEAT 2213 2243 NEBULIN 59.
 FI REPEAT 2248 2278 NEBULIN 60.
 FI REPEAT 2284 2314 NEBULIN 61.
 FI REPEAT 2322 2352 NEBULIN 62.
 FI REPEAT 2357 2387 NEBULIN 63.
 FI REPEAT 2418 2448 NEBULIN 64.
 FI REPEAT 2456 2486 NEBULIN 65.
 FI REPEAT 2491 2521 NEBULIN 66.
 FI REPEAT 2527 2557 NEBULIN 67.
 FI REPEAT 2565 2595 NEBULIN 68.
 FI REPEAT 2600 2630 NEBULIN 69.
 FI REPEAT 2631 2661 NEBULIN 70.
 FI REPEAT 2666 2696 NEBULIN 71.
 FI REPEAT 2699 2729 NEBULIN 72.
 FI REPEAT 2734 2764 NEBULIN 73.
 FI REPEAT 2770 2800 NEBULIN 74.
 FI REPEAT 2808 2838 NEBULIN 75.
 FI REPEAT 2843 2873 NEBULIN 76.
 FI REPEAT 2909 2939 NEBULIN 77.
 FI REPEAT 2942 2972 NEBULIN 78.
 FI REPEAT 2977 3007 NEBULIN 79.
 FI REPEAT 3013 3043 NEBULIN 80.
 FI REPEAT 3051 3081 NEBULIN 81.
 FI REPEAT 3086 3116 NEBULIN 82.
 FI REPEAT 3117 3147 NEBULIN 83.
 FI REPEAT 3152 3182 NEBULIN 84.
 FI REPEAT 3182 3212 NEBULIN 85.
 FI REPEAT 3212 3242 NEBULIN 86.

```

FT REPEAT 3185 3195 4-REP-11-17
FT REPEAT 3220 3230 4-REP-11-18
FT REPEAT 3256 3266 4-REP-11-19
FT REPEAT 3294 3304 4-REP-11-20
FT REPEAT 3329 3339 4-REP-11-21
FT REPEAT 3360 3370 4-REP-11-22
FT REPEAT 3395 3405 4-REP-11-23
FT REPEAT 3428 3438 4-REP-11-24
FT REPEAT 3463 3473 4-REP-11-25
FT REPEAT 3499 3509 4-REP-11-26
FT REPEAT 3537 3547 4-REP-11-27
FT REPEAT 3572 3582 4-REP-11-28
FT REPEAT 3603 3613 4-REP-11-29
FT REPEAT 3638 3648 4-REP-11-30
FT REPEAT 3671 3681 4-REP-11-31
FT REPEAT 3706 3716 4-REP-11-32
FT REPEAT 3742 3752 4-REP-11-33
FT REPEAT 3780 3790 4-REP-11-34
FT REPEAT 3815 3825 4-REP-11-35
FT REPEAT 3846 3856 4-REP-11-36
FT REPEAT 3894 3904 4-REP-11-37
FT REPEAT 3949 3959 4-REP-11-38
FT REPEAT 3984 3994 4-REP-11-39
FT REPEAT 4021 4031 4-REP-11-40
FT REPEAT 4057 4067 4-REP-11-41
FT REPEAT 4088 4098 4-REP-11-42
FT REPEAT 4123 4133 4-REP-11-43
FT REPEAT 4156 4166 4-REP-11-44
FT REPEAT 4191 4201 4-REP-11-45
FT REPEAT 4226 4236 4-REP-11-46
FT REPEAT 4264 4274 4-REP-11-47
FT REPEAT 4299 4309 4-REP-11-48
FT REPEAT 4340 4350 4-REP-11-49
FT REPEAT 4365 4375 4-REP-11-50
FT REPEAT 4400 4410 4-REP-11-51
FT REPEAT 4430 4440 4-REP-11-52
FT REPEAT 4471 4481 4-REP-11-53
FT REPEAT 4501 4511 4-REP-11-54
FT REPEAT 4544 4554 4-REP-11-55
FT REPEAT 4575 4585 4-REP-11-56
FT REPEAT 4610 4620 4-REP-11-57
FT REPEAT 4645 4655 4-REP-11-58
FT REPEAT 4680 4690 4-REP-11-59
FT REPEAT 4716 4726 4-REP-11-60
FT REPEAT 4754 4764 4-REP-11-61
FT REPEAT 4789 4799 4-REP-11-62

Query Match: 5119; Score: 11.1; Length: 5669,
Best Local Similarity: 22.2%; Id: 1; E: 1e-02
Matches: 46; Conservative: 24; Mismatches: 61; Indels: 76; Gaps: 9;

QY 20 QVLCRCHPARYVGGAKKQVSRERLV - KPMVVPVCHNLCPELPQATYR 69
DB 5076 QVLI-----HCYVVGITSDIKYKELQVLEKPELPYDTPMVRSHLR-----K 5119
QY 96 LAFNEQIEFLSMELPESFAYELVMVLAPEFAZELYSLEYLIMNEFVALYLANMEE 129
DB 5120 LNSN-----YIVDAKAWKAWKAVV-----LPIDYPAV -Q 5150
QY 140 LHEDKYKELFPWALDSHRLP - PQLPRKHPTWAKNRYAVYVRKCCDPMVMSKO 183
DB 5151 ELKTHLSLVYPAAKKFKSFTSVDTPELPAPFPEKTAQYLV -----PIATK- 5202
QY 194 LQAWALPDPHSHQZAMGSETPN -
DB 5263 ---EPHHHAJNQVIAIKKAKQ 531
RESULT 14
ID DPOZ_HUMAN STANDARD; ERT: 5.40 AA.
AC 060707.04214;
DT 15-DEC 1998 (Rel. 37, created)
DT 15-DEC 1998 (Rel. 37, Last sequence update)

```

```

11 C1-MAP-2002 (Rel. 41, Last annotation update)
12 DNA polymerase beta catalytic subunit (E: 2.7.7.7) (REV.)
13 REV3L OR POL2 OR REV3.
14 Homo sapiens (Human).
15 Pakaryeta, Metazoa: Chordata, Vertebrata, Euteleostomi:
16 Mammalia; Eutheria: Primates: Catarrhini; Hominidae; Homo.
17 NCBI_TaxID: 9606;
18 [1]
19 SEQUENCE FROM N.A.
20 MEDLINE: 98284025; PubMed: 9618506;
21 Gibbs P.E.M., McGregor W.G., Maher V.M., Nilsson P., Lawrence G.W.:
22 "A human homolog of the Saccharomyces cerevisiae REV3 gene, which
23 encodes the catalytic subunit of DNA polymerase beta."
24 Proc. Natl. Acad. Sci. U.S.A. 95:6876-6880(1998).
25 [2]
26 SEQUENCE FROM N.A.
27 TISSUE: Bone marrow, and Leukocyte;
28 MEDLINE: 99202263; PubMed: 10102035;
29 Lin W., Wu X., Wang Z.:
30 "A full-length cDNA of hREV3 is predicted to encode DNA polymerase
31 beta for damage induced mutagenesis in humans."
32 Mutat. Res. 433:89-98(1999).
33 [4]
34 SEQUENCE FROM N.A.
35 Murakumo Y., Rasio D., Roth T., Nearini M., Croce C.M., Finkel R.:
36 "Cloning and characterization of hREV3, the human homolog of S.
37 cerevisiae REV3."
38 Submitted (JUN 1999) to the EMBL/GenBank/DBJ databases.
39 [4]
40 SEQUENCE OF 79-3130 FROM N.A.
41 Roth T., Rasio D., Murakumo Y., Nearini M., Barbaudi-Brodano G., Croce C.M.:
42 "Alternative splicing, genomic structure, and fine chromosomal
43 localization of REV3."
44 Cytogenet. Cell Genet. 83:18-20(1998).
45 -- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
46 + (DNA)(N).
47 -- SUBCELLULAR LOCATION: Nuclear (potential).
48 -- TISSUE SPECIFICITY: UNICELLULARLY EXPRESSED.
49 -- DOMAIN: ITS C-TERMINAL PART COULD SERVE AS THE CATALYTIC DOMAIN
50 DURING NUCLEOTIDE POLYMERIZATION, WHILE ITS N-TERMINAL PART COULD
51 PROVIDE SITES FOR PROTEIN-PROTEIN INTERACTIONS WITH OTHER FACTORS
52 DURING TRANSLATION DNA SYNTHESIS.
53 -- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
54 -----
55 THIS SWISS-PROT entry is copyright. It is produced through a collaboration
56 between the Swiss Institute of Bioinformatics and the EMBL outstation -
57 the European Bioinformatics Institute. There are no restrictions on its
58 use by non-profit institutions as long as its content is in no way
59 modified and this statement is not removed. Usage by and for commercial
60 entities requires a license agreement (See http://www.ebi.ac.uk/seqdb/
61 or send an email to license@ebi.ac.uk).
62 -----
63 EMBL: AF058701; AAC24357.1;
64 EMBL: AF071798; AAC24009.1;
65 EMBL: AF157476; AAC40184.1;
66 EMBL: AF179428; AAC69402.1;
67 EMBL: AF179429; AAC09403.1;
68 EMBL: AF035537; AAC88486.1;
69 EMBL: AF078695; AAC28460.1;
70 MIM: 602776;
71 InterPro: IPR002064; DNA_pol_B.
72 Pfam: PF00136; DNA_pol_B; 1.
73 Pfam: PF03104; DNA_pol_H_exo; 2.
74 PRINTS: PR00106; DNAPOLH.
75 SMART: SM00486; POLRc; 1.
76 PROSITE: PS00116; DNA_POLYMERASE_B; 1.

```

Transferase; DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding; DNA repair; Nuclear protein; Zinc finger; Polymorphism.
 FI ZN_FING 3042 3057 C4 TYPE (POTENTIAL)
 FI ZN_FING 3086 3104 C4 TYPE (POTENTIAL)
 FI VARIANT 231 241 Q -> R
 FT FTID-VAR_008516.
 FT VARIANT 389 389 S -> T
 FT VARIANT 1540 1540 /FTID-VAR_008517.
 FT VARIANT 2607 2607 /FTID-VAR_008518.
 FT VARIANT 2607 2607 S -> T
 FT CONFLICT 247 247 E -> Q (IN REF. 4 AND 5).
 FT CONFLICT 1156 1156 Y -> C (IN REF. 4 AND 5).
 SQ SEQUENCE 3130 AA: 352782 MW: 100700900F10HR14 VR664.

Query Match
 Best Local Similarity 20.08; Pred. No: 62;
 Matches 62; Conservative 36; Mismatches 105; Indels 107; Gaps 13;

QY 31 VVCAKAOIPREELSVK PMVNRTHINIQPQEQAFYRIENQIQEPLSMISGLRIS 89
 DB 1577 VTSRKPRKSTKQKIPKLKAVSLNQ-----NSSQDNSVSSESIFPS 1625
 QY 90 DKYLIAMVLAYFKRANGLYTSEYTMFTFVALYLANDMEDEELYYEI - 148
 DB 1626 DP-----FFWA-IAGSWRELEFQLRLRDDFWAKNYIHAVSVRRCCDEVMSSKDT 1660
 QY 139 -----FFWA-IAGSWRELEFQLRLRDDFWAKNYIHAVSVRRCCDEVMSSKDT 185
 DB 1661 YSGSVEFVADQNLQPKELSDVAVQDFPGQATEKNEFLSHDNQK --CDEKRIHITD 1713
 QY 186 HWAMLR-----DRMHHSQAMGYLRNDDFFP---RGRPIIPASITLCHKAGVGD 243
 DB 1714 SASWIRSGTSPTEFKSTIDSNNRNQWKNKSFIPLTTRSNSIMDSFV--QNMEDCL 1771
 QY 234 SGGSVHNNSSSPEQEIF - - - - -HYTNREWSQEL - - - - -LMLPELL 270
 DB 1772 S-EKSLNRKSSVSREVFSLFQIPNNSWIOCHTKRMGOSLSANISFATLSSPPGCLV 1830
 QY 271 DPECTDLHI 280
 DB 1841 DVAG--EDLEL 1849

RESULT 15
 DIM_ARATH STANDARD; PRT: 561 AA.
 AC Q9085; Q9808;
 DI 15-JUL-1998 (Ref. 36, created)
 DI 01-MAR-2002 (Ref. 41, last sequence update)
 DI 01-MAR-2002 (Ref. 41, last annotation update)
 DE cell elongation protein DIMINUTO (cell elongation protein dwarf1).
 GN DIM OR DMF1 OR ATG19820 OR MPN9.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CV, COLUMBIA;
 RX MEDLINE=95129837; PubMed=7828854;
 RA Takahashi T., Gasch A., Nishizawa N., Chua N.-H.;
 RT "The DIMINUTO gene of Arabidopsis is involved in regulating cell
 RT elongation.";
 RL Genes Dev. 9:97-107(1995).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CV, COLUMBIA;
 RX Dukes B.P., Wu Y., Schulz B., Carlson T., Anna W., Feldman F.A.;
 RA Submitted (JUL-1994) to the EMBL/GenBank/Trna databases.

[4]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CV, COLUMBIA;
 RX MEDLINE=29277480; PubMed=10819429;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asanuma E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 4. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty B1 and 16
 RT clones.";
 RL DNA Res. 7:141-145(2000).
 CC -1- FUNCTION: PLAYS A CRITICAL ROLE IN THE GENERAL PROCESS OF PLANT
 CC CELL ELONGATION.
 CC -2- SIMILARITY: BELONGS TO THE DIMINUTO FAMILY.
 CC
 CC THIS SWISS-Prot entry is copyright . It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.ebi.ac.uk/edl/sa.html
 CC or send an email to license@isb.scrib).
 CC
 CC EMBL: L38520; AAA67055.1;
 CC EMBL: 012400; AAA20244.1;
 CC EMBL: AR025641; BA01296.1;
 CC InterPro: IPR001575; oxid_FAB bind.
 CC Pfam: PF01565; FAD binding_4; 1.
 CC CONFLICT 450 450 F -> G (IN REF. 1)
 CC CONFLICT 476 476 S -> L (IN REF. 1)
 CC CONFLICT 506 506 F -> L (IN REF. 1)
 CC CONFLICT 556 556 AYA -> PYP (IN REF. 1).
 SQ SEQUENCE 561 AA: 65094 MW: 70259795AM44299 VR664.

Query Match
 Best Local Similarity 5.08; Score 81.5; DB 1; Length 561;
 Matches 23; Conservative 14; Mismatches 16; Indels 27; Gaps 4;

QY 245 PEQEIFH - - - - -YTNREWSQEL MLEPPELL DEPT HIGH 279
 DB 341 PIRVYHHHTRCYWECKLLEPQDQFWNYLGLMLMPKVSLLKATGDAIKNYHMH 590
 QY 280 LLOEPLVLEPDTALEWHH 299
 DB 391 VIQMLVLYKYVGLAEWHV 410

Search completed: July 31, 2002, 17:11:42
 Job time: 254 sec


```

QY 121 LYLANMEDEDEEYKYEITFWALGDSWRELFPQFLRLRDPFWAKMNYKAVVSRRCQDEV 180
DB 121 LYLANMEDEDEEYKYEITFWALGDSWRELFPQFLRLRDPFWAKMNYKAVVSRRCQDEV 180
QY 181 SKDPTHWAWLRDRPMHHSKAMRYLYRNEDOFFPPGPGPLTPASLTTHKACVDSGVSHN 240
DB 181 SKDPTHWAWLRDRPMHHSKAMRYLYRNEDOFFPPGPGPLTPASLTTHKACVDSGVSHN 240
QY 241 NSSSQEIEFYTNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 300
DB 241 NSSSQEIEFYTNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 300
RESULT 2
QYFWR1
ID QYFWR1 PRELIMINARY: PKT: 299 AA.
AC QYFWR1
DT 01-MAY-2000 (Tremblrel, 14, Created)
DT 01-MAY-2000 (Tremblrel, 14, Last sequence update)
DT 01-JUN-2001 (Tremblrel, 17, Last annotation update)
DE P33 RINGO.
GN LS26/OLONE 5.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8455;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99396721; PubMed=10465793;
RA Forby L., Blazquez M., Palmer A., Britja R., Nebreda A.R.;
RT "A novel p34cdc2 binding and activator protein that is necessary and
RT sufficient to trigger G2/M progression in Xenopus oocytes."
RL Genes Dev. 13:2177-2189(1999).
DR EMBL: AJ249978; CAB58366.1;
DR INTERPRO: IPR000345; CYT home bind.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
SQ SEQUENCE 299 AA: 34660 MW: C4701F109KCAZ0E1 CR664;

```

```

Query Match 97.7% Score 1594.5; DB 13; Length 299;
Best Local Similarity 98.3% Pred. No. 9, 3e-150;
Matches 295; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

```

```

QY 1 MRHMVSSTRASSIGSGVKVIGKPHIAPVAVGAPKQIPEREELSVKPKMVENTHNLQ 60
DB 1 MRHMVSSTRASSIGSGVKVIGKPHIAPVAVGAPKQIPEREELSVKPKMVENTHNLQ 60
QY 61 PQRQAFYRLLENFQTOFFISMSCTKISIKYLIAVLAFFKAAIYISETYITMTTVA 120
DB 61 PQRQAFYRLLENFQTOFFISMSCTKISIKYLIAVLAFFKAAIYISETYITMTTVA 119
QY 121 LYLANMEDEDEEYKYEITFWALGDSWRELFPQFLRLRDPFWAKMNYKAVVSRRCQDEV 180
DB 121 LYLANMEDEDEEYKYEITFWALGDSWRELFPQFLRLRDPFWAKMNYKAVVSRRCQDEV 179
QY 181 SKDPTHWAWLRDRPMHHSKAMRYLYRNEDOFFPPGPGPLTPASLTTHKACVDSGVSHN 240
DB 181 SKDPTHWAWLRDRPMHHSKAMRYLYRNEDOFFPPGPGPLTPASLTTHKACVDSGVSHN 239
QY 241 NSSSQEIEFYTNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 300
DB 241 NSSSQEIEFYTNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 300
RESULT 3
QYFWR2
ID QYFWR2 PRELIMINARY: PKT: 298 AA.
AC QYFWR2
DT 01-MAY-2000 (Tremblrel, 14, Created)
DT 01-MAY-2000 (Tremblrel, 14, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE P33 RINGO.

```

```

GN LS27.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8455;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99396721; PubMed=10465793;
RA Forby L., Blazquez M., Palmer A., Britja R., Nebreda A.R.;
RT "A novel p34cdc2 binding and activator protein that is necessary and
RT sufficient to trigger G2/M progression in Xenopus oocytes."
RL Genes Dev. 13:2177-2189(1999).
DR EMBL: AJ143500; CAB44296.1;
SQ SEQUENCE 298 AA: 34573 MW: 88757D6C4278E26 CR664;

```

```

Query Match 89.5% Score 1460; DB 13; Length 298;
Best Local Similarity 89.0% Pred. No. 2e-136;
Matches 267; Conservative 15; Mismatches 19; Indels 2; Gaps 2;

```

```

QY 1 MRHMVSSTRASSIGSGVKVIGKPHIAPVAVGAPKQIPEREELSVKPKMVENTHNLQ 60
DB 1 MRHMVSSTRASSIGSGVKVIGKPHIAPVAVGAPKQIPEREELSVKPKMVENTHNLQ 60
QY 61 PQRQAFYRLLENFQTOFFISMSCTKISIKYLIAVLAFFKAAIYISETYITMTTVA 120
DB 61 PQRQAFYRLLENFQTOFFISMSCTKISIKYLIAVLAFFKAAIYISETYITMTTVA 119
QY 121 LYLANMEDEDEEYKYEITFWALGDSWRELFPQFLRLRDPFWAKMNYKAVVSRRCQDEV 180
DB 121 LYLANMEDEDEEYKYEITFWALGDSWRELFPQFLRLRDPFWAKMNYKAVVSRRCQDEV 179
QY 181 SKDPTHWAWLRDRPMHHSKAMRYLYRNEDOFFPPGPGPLTPASLTTHKACVDSGVSHN 240
DB 181 SKDPTHWAWLRDRPMHHSKAMRYLYRNEDOFFPPGPGPLTPASLTTHKACVDSGVSHN 239
QY 241 NSSSQEIEFYTNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 300
DB 241 NSSSQEIEFYTNREWSQELMLPPELLDPECTHDLHILOEPLVGLPDGTALFWHHL 298

```

```

RESULT 4
QYFWR1
ID QYFWR1 PRELIMINARY: PKT: 298 AA.
AC QYFWR1
DT 01-MAY-1999 (Tremblrel, 10, Created)
DT 01-MAY-1999 (Tremblrel, 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE SPEEDY PROTEIN.
GN SPY1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8455;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99219805; PubMed=10202150;
RA Leemans J.L., Deinum R.W., Knudsen K.E., Subramani N.;
RT "Speedy, a novel cyclin regulator of the G2/M transition."
RL EMBO J. 18:1869-1877(1999).
DR EMBL: AJ143117; CAB39117.1;
DR INTERPRO: IPR000345; CYC home bind.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
SQ SEQUENCE 298 AA: 34554 MW: 887531995F41A7 CR664;

```

```

Query Match 88.7% Score 1448; DB 13; Length 298;
Best Local Similarity 88.0% Pred. No. 6e-136;
Matches 264; Conservative 17; Mismatches 17; Indels 2; Gaps 2;

```

[illegible]

Query Match	44.9%	Score	276	nk 11	Length	241	
Best local Similarity	47.4%	Ident. No.	2,306-443				
Matches	111	Conservative	38	Misplaced	71	Indels	8
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
DB	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			
QY	1	MRHMSVTRASSGTSNGKAVIGKIKPHAVVAPAAQIPEETETAY	-----KPMVEN	54			

```

1 60  PCLTIGQEMIAAEKFKIDGAGAGQCEIWMVIGGCEKLAQVIGLAMIPTVYKRAKPTINRHTR 119
2 115  MRLIEVALYLANIMEEIEEESFYIEETWALIKSWEEELTILFLRLGLEWFMVSEAVVSR 174
3 120  INFFALYLANIVIEEIEEELAKKILIFPAWCKNNKRFLEFNAKLEGGWKEHLYPAIVSR 179
4 175  QDIEVNSKIDTHAWLIFKRIIMHSCAMKCYLIEHDFPTFQCLILFVASSLIL 226
5 180  QCEVMAIATPHYIWAPEKSVHHSZAVPNYNDE-VLLPFGFSATVDSLC 230
6
7 RESULT 6
8 QNTHS
9 PRELIMINARY: PPT: 308 AA.
10 QNTHS:
11 01-OCT-2000 (TREMBLrel. 15, Created)
12 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
13 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
14 BIOCHEMICAL 37.5 KDA PROTEIN (FRAGMENT).
15 DKE2P34A1014.
16 Homo sapiens (Human).
17 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
18 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
19 NCBI_TaxID=9606;
20 [?].
21 SEQUENCE FROM N.A.
22 TISSUE-TESTIS.
23 Kuehrer K., Meyer A., Meves H.W., Gassenhuber J., Wiemann S.;
24 Submitted (JAN 2000) to the EMBL/GenBank/TrnaJ databases.
25 EMBL: AL137266; CAH70665.1; -.
26 Hypothetical protein.
27 N-NTER 1
28 SEQUENCE 308 AA: 37491 MW: 0B79863A0B0E2817 CRC664.

```

```

Query Match. 19.0%; Score 309.5; BB 4; Length 308;
Best Local Similarity 34.9%; Pred. No. 2,1e+22;
Matches 75; Conservative 29; Mismatches 54; Indels 57; Gaps
27 35 PFACTPPEPES---SVETFWVRNTNINCPPEFAFYFVLEHFSQDELSMGQGLP 87
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
29 91 KKVLPAPPEELWVAEMCGGLMKLKKRPVSIVLPFHHEAFNRLLDDVVKRFLAWKDKLR 150
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
31 88 ISEYLIAMVATFYPAAGLYTSEYTIMNTVVALYLANMEESEFYETFTWALGDS 146
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
33 151 VSDRYLLAMVAFESP-AGEPSWYQPIHEFIALYLANMEEDEDSKONIEFLYKNE 209
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
35 147 -----WRKLFPO-----FLKLR 158
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
37 210 SRPILKRPFGILYSNMFAPAKNSHILVAFVEETN SEQMNPAPAKKSHIVLTFKE 269
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
39 159 DDFWAKMYPAVVSPPGCDDEVMSKDPHTWALWLEDP 193
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
41 370 FHPFCSKCAWVSPEELIEICAYDFEHWVWAWPER 304
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```

RESULT	7
Q9D5E0	
A / Q9D5E0	PRELIMINARY; PRT; 268 AA.
1	Q9D5E0:
2	01 JUN 2001 (TREMBLrel. 17, Created)
3	01 JUN-2001 (TREMBLrel. 17, Last sequence update)
4	01 JUN-2001 (TREMBLrel. 17, Last annotation update)
5	49A043LF05RK PROTEIN.
6	49A043LF05RK.
7	Mus musculus (Mouse).
8	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
9	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
X	NBI_TaxID=10090;
N	[1]
+	SEQUENCE FROM N.A.
+	STRAIN G57BL/76J; TISSUE=TESTIS;

RESULT	R
Q9D5G0	
ID	Q9D5G0
AC	PRELIMINARY; PRT; 173 AA.
DT	01-JUN-2001 (TEMBLrel. 17, Created)
DT	01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT	01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE	4930445A17RIK PROTEIN.
DE	4930445A17RIK
OS	Mus musculus (Mouse)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID	10090;
LN	111
RP	SEQUENCE FROM N.A.
SPRAIN	C57BL/6J; TISSUE=TESTIS;
EX	MEDLINE=21085660; PubMed=11217851;
EX	Kawai T., Shinagawa A., Shibata K., Yoshino M., Ishii Y.,
EX	Airawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
EX	Airawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakata I.,
EX	Saito T., Okazaki Y., Gojobori T., Hono H., Kasukawa T., Saito K.,
EX	Kadota K., Matsuo H.A., Ashburner M., Hatahov S., Sasavant T.,
EX	Fleischmann W., Gaasterland T., Gissi C., Kind T., Koriwa H.,
EX	Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pescio G., Quackenbush J.,
EX	Schriml L.M., Staubli F., Suzuki K., Tomita M., Walter L., Washio T.,
EX	Shake K., Okido T., Furuno M., Aono H., Baldarelli K., Barsh G.,
EX	Blake J., Roffelli D., Hojima N., Carninci P., de Remold M.F.,
EX	Brownstein M.J., Holt C., Fletcher C., Fujita M., Gariboldi M.,
EX	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
EX	Lyon P., Marchionni L., Mashima M., Mazzarelli J., Mombaerts P.,

Query Match	5.9%	Score 967	DB 107	Length 10087
Best Local Similarity	22.0%	Prod. No. 5		
Matches	55	Conservation 26	Mismatches 98	Indels 74
97	4	TSVTFWVWRTDINLQVTFZAFTHLDFLLETSMLTSTSTKYLIAVLAVIKFA	104	
DB	930	LSVKESALNVPQDLARRHTTF	ANSLEPMNMSAPICROMSESVLLPYKED	982
99	105	AGLTYSEYTMNFFVALYLANIMEEED	YKVEYFWALQESW	147
DB	983	-----GVLYSEELNRENELGSLFYQKTYF	LEWNYLRLRGLR	107
99	138	PELEPQPLDQGGGWSK	MNRYAVVSGPCTFYVSKETLHWALRQTHD	190
DB	1024	LPEKIKSETEWVSVYRQLAFIVPGWVYQALQLQVCEV	ASLQALHNVIF	107
99	137	HS-AMRGLRLRHLRFLFRGLFLPASFL	CUKAVTFRQVSVNNRSETEQLQTHPT	253
DB	1078	AMASNTQNKAFTEKALADLKEIYVVSQVYVQNKKSGLHNRG		1129
99	254	N 254		

01-MAY-2000 (TRENBLrel. 13, Created)
01-MAY 2000 (TRENBLrel. 13, Last sequence update)
01-DEC-2001 (TRENBLrel. 19, Last annotation update)
CG9747 PROTEIN (GH07782P).
CG9747.
Drosophila melanogaster (Fruit fly).
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscophora;
Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
STRAIN=BERKELEY.
MELTINE-201606; PubMed:10731132;
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Curtis G.W., Wortman W.P., Yu J., Zeng Q., Chen L.X.,
Brandon R.C., Rogers J., Haussler R.G., Champ M.P., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson G.R., Miklos G.L.G.,
Abbitt J.P., Achwayan A., An H. J., Andrews-Pfankuch C., Baldwin D.,
Balew F.M., Bass A., Baxendale J., Bayraktaroglu L., Becksky E.M.,
Beeson K.Y., Bonos P.V., Borman H.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.P., Brock J., Bruck E., Brutsaert P.,
Buttis K.C., Busch D.A., Butler H., Cadien E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablo B., Deitcher A., Deng Z., Mays A.B., Dew I., Bier S.M.,
Peterson K., Doup E.K., Dowes M., Egan-Kocha S., Bankov H.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Fowler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Hartis M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.P., Hoyle J.,
Rustin D., Houston K.A., Howland T.J., Wei M.-H., Iqbal M.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kotchum K.A.,
Kucinski H.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Lian Y., Lin X.,
Liu X., Mattie B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkushev G., Milshina N.V., McBarry C., Morris J., Mochizuki A.,
Mount S.M., Moy M., Murphy B., Murphy L., Murny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacheco J.M.,
Palazzo M., Pittman G.S., Pan S., Pelland J., Puri V., Reese M.G.,
Reiner K., Remington K., Saunders P.D., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
The genome sequence of *Drosophila melanogaster*.;
Science 287:2185-2195(2009).
[2]
SEQUENCE FROM N.A.
STRAIN=Y. CN BW SP;
Stapleton M., Brokstein P., Hong L., Achyani A., Carlson J.,
Champe M., Charvet G., Lerest V., Rafan D., Frise E., George R.,
Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mundall J.,
Munoz J., Paclet J., Paragay V., Park S., Pheasant-Craig G., Wan K.,
Yu C., Lewis S.E., Rubin G.M., Celniker S.,
Submitted (URL: 557), to the EMBL/GenBank/DBI databases.
EMBL: AE003772; AAF57020.1;
FlyBase: AY047545; AAK7297.1;
FlyBase: FBgn0039754; CG9747.
InterPro: IPR001522; Desaturase.
Pfam: PF01069; Desaturase_1
PRINIS: PR000075; FADDSATURASE.
ProDom: PD002221; Desaturase; 1.
SEQUENCE 461 AA; 53634 MW; 3271892930285FF CRC64;
Query Match 5.78; Score 94; DB 5; Length 461;
Best Local Similarity 23.08; Ered. No. 1;

Matches 68: Conservative 46; Mismatches 104; Indels 88; Gaps 16;

QY 7 VTRASSICGSGYKQVIGKHGHARVVGAR-KAGI-PKREEL SVKKMKVNTNHNLIQPO 52
 DB 26 VAKRSKPPNCDAQOQVRKRRNSPELPYCKIOLPPKRELLRSVVIKELIAN HVPKAI 84
 QY 63 ERQAFYRLLENE-----QIQEFLSMDSCLRSIDKYLIA--MVLAYFKKAAGLYTSYTTM 115
 DB 85 DUEKYTRTFVGDRVIGWQFOAPLKWDKVIOISLHIVAGICILYHDLRELNYT---TIV 141
 QY 116 NFEVALYIANDMEDEEDYKYEFIFWALGDSWRELFPQFLRIKDPWAKMNYHRAVVSER- 174
 DB 142 STFGGVAG-----FVTAQ-----AHPWTHRSYKANTYLSR 174
 QY 175 -CQDEVMSKQPIHWAMLRDPMHHSAMCYLRNDDFFPQGLIIPASCTLCHKAGV 241
 DB 175 ELMVSYVAGQN-ELYDWRVDRVHRHK-----YSETDADE----- 208
 QY 232 CDSGGVSHNNSSSPQEIFHYTNREWSQELLMPL-PELLLDPECTHMLHLLQELPLV 286
 DB 209 -HNANKG-----FFSHVGW---LMLKHPDVLRRG-ROTUMSDHIAADVV 249

RESULT 12
 Q91AW0
 ID Q91AW0 PRELIMINARY: PRT: 1080 AA.
 AC Q91AW0:
 DT 01-DEC-2001 (Tremblrel, 19, Created)
 DT 01-DEC-2001 (Tremblrel, 19, Last sequence update)
 DE PUTATIVE FUSION PROTEIN P1 P2.
 OS beel chlorosis virus.
 OC Viruses; ssRNA positive strand viruses, no DNA stage; Lutovoviridae;
 OC Polioviruses.
 OC NCBI_TaxID=131082;
 RN [1]
 RC STRAIN:BCIV-2A.
 RA Lemaitre G., Hauser S., Stevens M., Heve M.;
 RT "Biological properties and molecular characterization of beel
 chlorosis virus (BCIV).";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF452024; AAK49956.1;
 SQ SEQUENCE 1080 AA: 121612 MW: 22780.400 kDa

Query Match
 Best Local Similarity 22.7%; Score 92; DB 12; Length 1080;
 Matches 72; Conservative 41; Mismatches 138; Indels 66; Gaps 16;

QY 11 SSICG-----SGYKQVIGKCHPHA-----RVVGAKKA QIPKRELSVKIKMYRNTH 56
 DB 341 ASUSAYIKKASVSTIDGHSAPYISCKNVGCHSGGVHNVGSLPKIVNLIIT 400
 QY 57 NLQPKQAFYRLLENEQIQEFLSMDSCLRSIDKYLIAWVLAYFKKAAGLYTSYTTM 116
 DB 401 POLVFETTPQGRFLKEMDILLIYD-----WSQARALVKEPKE-GLYFGKGVK--- 452
 QY 117 FEVALYIANDMEDEEDYKYEFIFWALGDSWRELFPQFLRIKDPWAKMNYHRAVVSER- 171
 DB 453 -----SYFETTSQVSVPIVPPHLPKWKTERPPPPNNPHIFH-PEKSNRLWPELK 503
 QY 172 SRGCTGVMSKQPIHWAMLRDPMHHSAMCYLRNDDFFPQGLIIPASCTLCHKAGV 241
 DB 504 ARSCUDS-AKPFTHNASW-----NTDTPDPRCTHDSHTCTFCGS 544
 QY 226 CHKAGVDSGVSNNSSSPQEIFHYTNREWSQELLMPL-PELLLDPECTHMLHLLQELPLV 282
 DB 545 KHESCI-DCSHPFNKSGREYRE SCWTDSEKREIAIKTQAKLEKFEFGYHFWKSCHTT 602
 QY 283 EPLWGLDGTALPMWH 299
 DB 603 QEVGTEEVGSLPKFYH 619

RESULT 14
 Q5B530
 ID Q5B530 PRELIMINARY: PRT: 1624 AA.
 AC Q5B530:
 DT 01-AUG-1998 (Tremblrel, 07, Created)
 DT 01-AUG-1998 (Tremblrel, 07, Last sequence update)
 DE 1624AA LONG HYPOTHETICAL REVERSE GYRASE.
 GN PH0800.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.
 OC NCBI_TaxID=54554;
 RN 111
 RC SEQUENCE FROM N.A.
 RC STRAIN 014;
 RX MF01NF-08444137; PubMed 9679194;
 RA Kawarabayashi Y., Sawada M., Horikawa H., Baikawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baba S., Ito K., Kusunagi H., Bessyama A., Nagai Y.,
 RA Sakai M., Ogura K., Otsuki R., Nakazawa H., Takamiya M., Ohfuku Y.,
 RA Fushibashi T., Tanaka T., Yoshizawa T., Yamazaki J., Koshida N., Tanuchi A.,
 RA Aoki K., Ito Y., Shizuya H., Nakamura Y., Robb F.L., Horikoshi K.,
 RA Masuchi Y., Shizuya H., Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 thermophilic archaeobacterium, Pyrococcus horikoshii, of 4.3";
 RL DNA Res. 5:55-76(1998).
 DR EMBL: AP000004; XAA29894.1;
 DR HSSU: 006612; 1871;
 DR InterPro: IPR001410; IPR001410;
 DR InterPro: IPR003601; IPR003601; IPR003601; IPR003601;
 DR InterPro: IPR003602; IPR003602; IPR003602; IPR003602;
 DR InterPro: IPR003603; IPR003603; IPR003603; IPR003603;
 DR InterPro: IPR003604; IPR003604; IPR003604; IPR003604;
 DR InterPro: IPR003605; IPR003605; IPR003605; IPR003605;
 DR InterPro: IPR003606; IPR003606; IPR003606; IPR003606;
 DR InterPro: IPR003607; IPR003607; IPR003607; IPR003607;
 DR InterPro: IPR003608; IPR003608; IPR003608; IPR003608;
 DR InterPro: IPR003609; IPR003609; IPR003609; IPR003609;
 DR InterPro: IPR003610; IPR003610; IPR003610; IPR003610;
 DR InterPro: IPR003611; IPR003611; IPR003611; IPR003611;
 DR InterPro: IPR003612; IPR003612; IPR003612; IPR003612;
 DR InterPro: IPR003613; IPR003613; IPR003613; IPR003613;
 DR InterPro: IPR003614; IPR003614; IPR003614; IPR003614;
 DR InterPro: IPR003615; IPR003615; IPR003615; IPR003615;
 DR InterPro: IPR003616; IPR003616; IPR003616; IPR003616;
 DR InterPro: IPR003617; IPR003617; IPR003617; IPR003617;
 DR InterPro: IPR003618; IPR003618; IPR003618; IPR003618;
 DR InterPro: IPR003619; IPR003619; IPR003619; IPR003619;
 DR InterPro: IPR003620; IPR003620; IPR003620; IPR003620;
 DR InterPro: IPR003621; IPR003621; IPR003621; IPR003621;
 DR InterPro: IPR003622; IPR003622; IPR003622; IPR003622;
 DR InterPro: IPR003623; IPR003623; IPR003623; IPR003623;
 DR InterPro: IPR003624; IPR003624; IPR003624; IPR003624;
 DR InterPro: IPR003625; IPR003625; IPR003625; IPR003625;
 DR InterPro: IPR003626; IPR003626; IPR003626; IPR003626;
 DR InterPro: IPR003627; IPR003627; IPR003627; IPR003627;
 DR InterPro: IPR003628; IPR003628; IPR003628; IPR003628;
 DR InterPro: IPR003629; IPR003629; IPR003629; IPR003629;
 DR InterPro: IPR003630; IPR003630; IPR003630; IPR003630;
 DR InterPro: IPR003631; IPR003631; IPR003631; IPR003631;
 DR InterPro: IPR003632; IPR003632; IPR003632; IPR003632;
 DR InterPro: IPR003633; IPR003633; IPR003633; IPR003633;
 DR InterPro: IPR003634; IPR003634; IPR003634; IPR003634;
 DR InterPro: IPR003635; IPR003635; IPR003635; IPR003635;
 DR InterPro: IPR003636; IPR003636; IPR003636; IPR003636;
 DR InterPro: IPR003637; IPR003637; IPR003637; IPR003637;
 DR InterPro: IPR003638; IPR003638; IPR003638; IPR003638;
 DR InterPro: IPR003639; IPR003639; IPR003639; IPR003639;
 DR InterPro: IPR003640; IPR003640; IPR003640; IPR003640;
 DR InterPro: IPR003641; IPR003641; IPR003641; IPR003641;
 DR InterPro: IPR003642; IPR003642; IPR003642; IPR003642;
 DR InterPro: IPR003643; IPR003643; IPR003643; IPR003643;
 DR InterPro: IPR003644; IPR003644; IPR003644; IPR003644;
 DR InterPro: IPR003645; IPR003645; IPR003645; IPR003645;
 DR InterPro: IPR003646; IPR003646; IPR003646; IPR003646;
 DR InterPro: IPR003647; IPR003647; IPR003647; IPR003647;
 DR InterPro: IPR003648; IPR003648; IPR003648; IPR003648;
 DR InterPro: IPR003649; IPR003649; IPR003649; IPR003649;
 DR InterPro: IPR003650; IPR003650; IPR003650; IPR003650;
 DR InterPro: IPR003651; IPR003651; IPR003651; IPR003651;
 DR InterPro: IPR003652; IPR003652; IPR003652; IPR003652;
 DR InterPro: IPR003653; IPR003653; IPR003653; IPR003653;
 DR InterPro: IPR003654; IPR003654; IPR003654; IPR003654;
 DR InterPro: IPR003655; IPR003655; IPR003655; IPR003655;
 DR InterPro: IPR003656; IPR003656; IPR003656; IPR003656;
 DR InterPro: IPR003657; IPR003657; IPR003657; IPR003657;
 DR InterPro: IPR003658; IPR003658; IPR003658; IPR003658;
 DR InterPro: IPR003659; IPR003659; IPR003659; IPR003659;
 DR InterPro: IPR003660; IPR003660; IPR003660; IPR003660;
 DR InterPro: IPR003661; IPR003661; IPR003661; IPR003661;
 DR InterPro: IPR003662; IPR003662; IPR003662; IPR003662;
 DR InterPro: IPR003663; IPR003663; IPR003663; IPR003663;
 DR InterPro: IPR003664; IPR003664; IPR003664; IPR003664;
 DR InterPro: IPR003665; IPR003665; IPR003665; IPR003665;
 DR InterPro: IPR003666; IPR003666; IPR003666; IPR003666;
 DR InterPro: IPR003667; IPR003667; IPR003667; IPR003667;
 DR InterPro: IPR003668; IPR003668; IPR003668; IPR003668;
 DR InterPro: IPR003669; IPR003669; IPR003669; IPR003669;
 DR InterPro: IPR003670; IPR003670; IPR003670; IPR003670;
 DR InterPro: IPR003671; IPR003671; IPR003671; IPR003671;
 DR InterPro: IPR003672; IPR003672; IPR003672; IPR003672;
 DR InterPro: IPR003673; IPR003673; IPR003673; IPR003673;
 DR InterPro: IPR003674; IPR003674; IPR003674; IPR003674;
 DR InterPro: IPR003675; IPR003675; IPR003675; IPR003675;
 DR InterPro: IPR003676; IPR003676; IPR003676; IPR003676;
 DR InterPro: IPR003677; IPR003677; IPR003677; IPR003677;
 DR InterPro: IPR003678; IPR003678; IPR003678; IPR003678;
 DR InterPro: IPR003679; IPR003679; IPR003679; IPR003679;
 DR InterPro: IPR003680; IPR003680; IPR003680; IPR003680;
 DR InterPro: IPR003681; IPR003681; IPR003681; IPR003681;
 DR InterPro: IPR003682; IPR003682; IPR003682; IPR003682;
 DR InterPro: IPR003683; IPR003683; IPR003683; IPR003683;
 DR InterPro: IPR003684; IPR003684; IPR003684; IPR003684;
 DR InterPro: IPR003685; IPR003685; IPR003685; IPR003685;
 DR InterPro: IPR003686; IPR003686; IPR003686; IPR003686;
 DR InterPro: IPR003687; IPR003687; IPR003687; IPR003687;
 DR InterPro: IPR003688; IPR003688; IPR003688; IPR003688;
 DR InterPro: IPR003689; IPR003689; IPR003689; IPR003689;
 DR InterPro: IPR003690; IPR003690; IPR003690; IPR003690;
 DR InterPro: IPR003691; IPR003691; IPR003691; IPR003691;
 DR InterPro: IPR003692; IPR003692; IPR003692; IPR003692;
 DR InterPro: IPR003693; IPR003693; IPR003693; IPR003693;
 DR InterPro: IPR003694; IPR003694; IPR003694; IPR003694;
 DR InterPro: IPR003695; IPR003695; IPR003695; IPR003695;
 DR InterPro: IPR003696; IPR003696; IPR003696; IPR003696;
 DR InterPro: IPR003697; IPR003697; IPR003697; IPR003697;
 DR InterPro: IPR003698; IPR003698; IPR003698; IPR003698;
 DR InterPro: IPR003699; IPR003699; IPR003699; IPR003699;
 DR InterPro: IPR003700; IPR003700; IPR003700; IPR003700;
 DR InterPro: IPR003701; IPR003701; IPR003701; IPR003701;
 DR InterPro: IPR003702; IPR003702; IPR003702; IPR003702;
 DR InterPro: IPR003703; IPR003703; IPR003703; IPR003703;
 DR InterPro: IPR003704; IPR003704; IPR003704; IPR003704;
 DR InterPro: IPR003705; IPR003705; IPR003705; IPR003705;
 DR InterPro: IPR003706; IPR003706; IPR003706; IPR003706;
 DR InterPro: IPR003707; IPR003707; IPR003707; IPR003707;
 DR InterPro: IPR003708; IPR003708; IPR003708; IPR003708;
 DR InterPro: IPR003709; IPR003709; IPR003709; IPR003709;
 DR InterPro: IPR003710; IPR003710; IPR003710; IPR003710;
 DR InterPro: IPR003711; IPR003711; IPR003711; IPR003711;
 DR InterPro: IPR003712; IPR003712; IPR003712; IPR003712;
 DR InterPro: IPR003713; IPR003713; IPR003713; IPR003713;
 DR InterPro: IPR003714; IPR003714; IPR003714; IPR003714;
 DR InterPro: IPR003715; IPR003715; IPR003715; IPR003715;
 DR InterPro: IPR003716; IPR003716; IPR003716; IPR003716;
 DR InterPro: IPR003717; IPR003717; IPR003717; IPR003717;
 DR InterPro: IPR003718; IPR003718; IPR003718; IPR003718;
 DR InterPro: IPR003719; IPR003719; IPR003719; IPR003719;
 DR InterPro: IPR003720; IPR003720; IPR003720; IPR003720;
 DR InterPro: IPR003721; IPR003721; IPR003721; IPR003721;
 DR InterPro: IPR003722; IPR003722; IPR003722; IPR003722;
 DR InterPro: IPR003723; IPR003723; IPR003723; IPR003723;
 DR InterPro: IPR003724; IPR003724; IPR003724; IPR003724;
 DR InterPro: IPR003725; IPR003725; IPR003725; IPR003725;
 DR InterPro: IPR003726; IPR003726; IPR003726; IPR003726;
 DR InterPro: IPR003727; IPR003727; IPR003727; IPR003727;
 DR InterPro: IPR003728; IPR003728; IPR003728; IPR003728;
 DR InterPro: IPR003729; IPR003729; IPR003729; IPR003729;
 DR InterPro: IPR003730; IPR003730; IPR003730; IPR003730;
 DR InterPro: IPR003731; IPR003731; IPR003731; IPR003731;
 DR InterPro: IPR003732; IPR003732; IPR003732; IPR003732;
 DR InterPro: IPR003733; IPR003733; IPR003733; IPR003733;
 DR InterPro: IPR003734; IPR003734; IPR003734; IPR003734;
 DR InterPro: IPR003735; IPR003735; IPR003735; IPR003735;
 DR InterPro: IPR003736; IPR003736; IPR003736; IPR003736;
 DR InterPro: IPR003737; IPR003737; IPR003737; IPR003737;
 DR InterPro: IPR003738; IPR003738; IPR003738; IPR003738;
 DR InterPro: IPR003739; IPR003739; IPR003739; IPR003739;
 DR InterPro: IPR003740; IPR003740; IPR003740; IPR003740;
 DR InterPro: IPR003741; IPR003741; IPR003741; IPR003741;
 DR InterPro: IPR003742; IPR003742; IPR003742; IPR003742;
 DR InterPro: IPR003743; IPR003743; IPR003743; IPR003743;
 DR InterPro: IPR003744; IPR003744; IPR003744; IPR003744;
 DR InterPro: IPR003745; IPR003745; IPR003745; IPR003745;
 DR InterPro: IPR003746; IPR003746; IPR003746; IPR003746;
 DR InterPro: IPR003747; IPR003747; IPR003747; IPR003747;
 DR InterPro: IPR003748; IPR003748; IPR003748; IPR003748;
 DR InterPro: IPR003749; IPR003749; IPR003749; IPR003749;
 DR InterPro: IPR003750; IPR003750; IPR003750; IPR003750;
 DR InterPro: IPR003751; IPR003751; IPR003751; IPR003751;
 DR InterPro: IPR003752; IPR003752; IPR003752; IPR003752;
 DR InterPro: IPR003753; IPR003753; IPR003753; IPR003753;
 DR InterPro: IPR003754; IPR003754; IPR003754; IPR003754;
 DR InterPro: IPR003755; IPR003755; IPR003755; IPR003755;
 DR InterPro: IPR003756; IPR003756; IPR003756; IPR003756;
 DR InterPro: IPR003757; IPR003757; IPR003757; IPR003757;
 DR InterPro: IPR003758; IPR003758; IPR003758; IPR003758;
 DR InterPro: IPR003759; IPR003759; IPR003759; IPR003759;
 DR InterPro: IPR003760; IPR003760; IPR003760; IPR003760;
 DR InterPro: IPR003761; IPR003761; IPR003761; IPR003761;
 DR InterPro: IPR003762; IPR003762; IPR003762; IPR003762;
 DR InterPro: IPR003763; IPR003763; IPR003763; IPR003763;
 DR InterPro: IPR003764; IPR003764; IPR003764; IPR003764;
 DR InterPro: IPR003765; IPR003765; IPR003765; IPR003765;
 DR InterPro: IPR003766; IPR003766; IPR003766; IPR003766;
 DR InterPro: IPR003767; IPR003767; IPR003767; IPR003767;
 DR InterPro: IPR003768; IPR003768; IPR003768; IPR003768;
 DR InterPro: IPR003769; IPR003769; IPR003769; IPR003769;
 DR InterPro: IPR003770; IPR003770; IPR003770; IPR003770;
 DR InterPro: IPR003771; IPR003771; IPR003771; IPR003771;
 DR InterPro: IPR003772; IPR003772; IPR003772; IPR003772;
 DR InterPro: IPR003773; IPR003773; IPR003773; IPR003773;
 DR InterPro: IPR003774; IPR003774; IPR003774; IPR003774;
 DR InterPro: IPR003775; IPR003775; IPR003775; IPR003775;
 DR InterPro: IPR003776; IPR003776; IPR003776; IPR003776;
 DR InterPro: IPR003777; IPR003777; IPR003777; IPR003777;
 DR InterPro: IPR003778; IPR003778; IPR003778; IPR003778;
 DR InterPro: IPR003779; IPR003779; IPR003779; IPR003779;
 DR InterPro: IPR003780; IPR003780; IPR003780; IPR003780;
 DR InterPro: IPR003781; IPR003781; IPR003781; IPR003781;
 DR InterPro: IPR003782; IPR003782; IPR003782; IPR003782;
 DR InterPro: IPR003783; IPR003783; IPR003783; IPR003783;
 DR InterPro: IPR003784; IPR003784; IPR003784; IPR003784;
 DR InterPro: IPR003785; IPR003785; IPR003785; IPR003785;
 DR InterPro: IPR003786; IPR003786; IPR003786; IPR003786;
 DR InterPro: IPR003787; IPR003787; IPR003787; IPR003787;
 DR InterPro: IPR003788; IPR003788; IPR003788; IPR003788;
 DR InterPro: IPR003789; IPR003789; IPR003789; IPR003789;
 DR InterPro: IPR003790; IPR003790; IPR003790; IPR003790;
 DR InterPro: IPR003791; IPR003791; IPR003791; IPR003791;
 DR InterPro: IPR003792; IPR003792; IPR003792; IPR003792;
 DR InterPro: IPR003793; IPR003793; IPR003793; IPR003793;
 DR InterPro: IPR003794; IPR003794; IPR003794; IPR003794;
 DR InterPro: IPR003795; IPR003795; IPR003795; IPR003795;
 DR InterPro: IPR003796; IPR003796; IPR003796; IPR003796;
 DR InterPro: IPR003797; IPR003797; IPR003797; IPR003797;
 DR InterPro: IPR003798; IPR003798; IPR003798; IPR003798;
 DR InterPro: IPR003799; IPR003799; IPR003799; IPR003799;
 DR InterPro: IPR003800; IPR003800; IPR003800; IPR003800;
 DR InterPro: IPR003801; IPR003801; IPR003801; IPR003801;
 DR InterPro: IPR003802; IPR003802; IPR003802; IPR003802;
 DR InterPro: IPR003803; IPR003803; IPR003803; IPR003803;
 DR InterPro: IPR003804; IPR003804; IPR003804; IPR003804;
 DR InterPro: IPR003805; IPR003805; IPR003805; IPR003805;
 DR InterPro: IPR003806; IPR003806; IPR003806; IPR003806;
 DR InterPro: IPR003807; IPR003807; IPR003807; IPR003807;
 DR InterPro: IPR003808; IPR003808; IPR003808; IPR003808;
 DR InterPro: IPR003809; IPR003809; IPR003809; IPR003809;
 DR InterPro: IPR003810; IPR003810; IPR003810; IPR003810;
 DR InterPro: IPR003811; IPR003811; IPR003811; IPR003811;
 DR InterPro: IPR003812; IPR003812; IPR003812; IPR003812;
 DR InterPro: IPR003813; IPR003813; IPR003813; IPR003813;
 DR InterPro: IPR003814; IPR003814; IPR003814; IPR003814;
 DR InterPro: IPR003815; IPR003815; IPR003815; IPR003815;
 DR InterPro: IPR003816; IPR003816; IPR003816; IPR003816;
 DR InterPro: IPR003817; IPR003817; IPR003817; IPR003817;
 DR InterPro: IPR003818; IPR003818; IPR003818; IPR003818;
 DR InterPro: IPR003819; IPR003819; IPR003819; IPR003819;
 DR InterPro: IPR003820; IPR003820; IPR003820; IPR003820;
 DR InterPro: IPR003821; IPR003821; IPR003821; IPR003821;
 DR InterPro: IPR003822; IPR003822; IPR003822; IPR003822;
 DR InterPro: IPR003823; IPR003823; IPR003823; IPR003823;
 DR InterPro: IPR003824; IPR003824; IPR003824; IPR003824;
 DR InterPro: IPR003825; IPR003825; IPR003825; IPR003825;
 DR InterPro: IPR003826; IPR003826; IPR003826; IPR003826;
 DR InterPro: IPR003827; IPR003827; IPR003827; IPR003827;
 DR InterPro: IPR003828; IPR003828; IPR003828; IPR003828;
 DR InterPro: IPR003829; IPR003829; IPR003829; IPR003829;
 DR InterPro: IPR003830; IPR003830; IPR0038

DB 1403 KSKADFLONEPQDQAKKAYESIGCLDAIPGVNPMKSRZLFTAGSGSNLSIIHLESJHLD 1462
QY 221 -ASRTIC--- HKACVDSGCVSHNNSSSDQEHFHYTNKRWSCQELIMPPPELLLDPECT 275
DB 1463 KASGQCIDIMKGNGLNSFAKLQJQJQNEPDYEIF-WRLQW-- -DSLIDPKHQ 1512
QY 276 HJLHLLQEPLVGLPEPDGTALEMHH 299
DB 1513 ONOTVVRTSU-DLEQE- -FKRHH 1532

Search completed: July 31, 2002, 17:10:51
Job time: 268 sec

A:Status: translated from cDNA, EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-17 <RES>
 A:Cross-references: GH:M58478; NID:q180291; PDB:AAFG1942.1; PDB:180292
 A:Note: neither the complete nucleic acid sequence nor the complete translation are shown
 C:Comment: This protein (CFTR) is directly responsible for cystic fibrosis (CF), a disor-
 plasia of the exocrine glands. A single residue deletion (Phe 508) is detected in most CF patients; how-
 ever, this mutation is not found in the apical membrane of epithelial cells
 C:Genetics:
 A:Gene: CFTR; CF
 A:Cross-references: GH:M58478; NID:q180291; PDB:AAFG1942.1; PDB:180292
 A:Map position: 7q31.2-7q31.2
 A:Introns: 18/2; Exons: 13/3
 C:Superfamily: cystic fibrosis transmembrane conductance regulator; ATP-binding cassette
 C:Keywords: ATP; duplication; transmembrane; nucleotide-binding; P-loop; phosphoprotein;
 F:81-102/domain: transmembrane #status predicted <IM>
 F:118-148/domain: transmembrane #status predicted <IM>
 F:195-215/domain: transmembrane #status predicted <IM>
 F:221-241/domain: transmembrane #status predicted <IM>
 F:308-328/domain: transmembrane #status predicted <IM>
 F:331-350/domain: transmembrane #status predicted <IM>
 F:441-622/domain: ATP-binding cassette homology <ABC>
 F:568-572/region: nucleotide-binding motif A (P-loop)
 F:593-840/region: nucleotide-binding motif B
 F:860-880/domain: transmembrane #status predicted <IM>
 F:912-932/domain: transmembrane #status predicted <IM>
 F:991-1011/domain: transmembrane #status predicted <IM>
 F:1103-1123/domain: transmembrane #status predicted <IM>
 F:1129-1150/domain: transmembrane #status predicted <IM>
 F:1227-1419/domain: ATP binding cassette homology <ABC>
 F:1244-1251/region: nucleotide-binding motif A (P-loop)
 F:1366-1370/region: nucleotide-binding motif B
 F:1464/binding site: ATP (lys) #status predicted
 F:660,700,737,813/binding site: phosphate (Ser) (covalent)
 F:686,790/binding site: phosphate (Ser) (covalent)
 F:768,795/binding site: phosphate (Ser) (covalent)
 F:894,900/binding site: carbohydrate (Asn) (covalent)
 F:1250/binding site: ATP (lys) #status predicted

Query Match 5.4% Score 86; DB 1; Length 1480;
 Best Local Similarity 20.7% Pred. No. 46;
 Matches 42; Conservative 47; Mismatches 74; Indels 50; Gaps 8;
 QY 9 RASSICSGGVKVI -- GKGHFARVGVKCAKQIPERELSVKPKVKNTHLQIQE 63
 Db 220 QANAFGLGLFLVLFQALGRLMMKYPDQPAKISER--LVITSEMENIQSVKAVCW 277
 QY 64 RQAFYRLLENEQTEFLSMNSGLRISDKYLIAMVLAYTKRAARLYT----- 109
 Db 278 FEAMKMIENLRQTE-----LKLIRK---AAVYVYFNSSAFESGFVVFSLGVYAL 327
 QY 110 -----SEYTMNFFVALYLANDMEDEEDYKYLFPWALGDSWRELFPDFLRDRFPW 162
 Db 328 IKGILIRKIFTTISFICVLRMAVTRQ-----FPWAV-UTWYSLGAINKIQ-DEL 375
 QY 163 AKMYRAVYVSRRCDFVMSKDPH 185
 Db 376 QKQYKTLVNLTTTEVVMNT 498
 RESULT 14
 AC2091
 serine/threonine kinase with two-component sensor domain; a12282 [unpublished] - Anabaena s-
 C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PVA 7120) is a synonym of Nostoc sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #taxi_change 11 Jan 2002
 C:Accession: AC2091
 K.Kaneko, I.; Nakamura, Y.; Wolk, C.P.; Kunitz, F.; Sasamoto, S.; Watanabe, A.; Iriyuchi,
 N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-214, 2001
 A:Title: Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium An-

A:Reference number: AB1807; MIM:21595280; PDB:1175040
 A:Accession: AC2091
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1783 <KOR>
 A:Cross-references: GH:BA000019; PDB:BAW7081.1; PDB:1714374; GDB:G000179
 A:Experimental source: strain per 7120
 C:Genetics:
 A:Gene: a12282

Query Match 5.2% Score 85.5; DB 2; Length 1783;
 Best Local Similarity 20.2% Pred. No. 51;
 Matches 46; Conservative 46; Mismatches 81; Indels 65; Gaps 9;
 QY 36 KAGIPERELSVKPKVKNTHLQIQEGRFVY ELLENQTEFLSMNSGLRISDKYLI 94
 Db 1143 KGLPVALGQNYQSLLHASTVYKVTPEQAFWQFENNEFMTWAINSEFLIKYL 1201
 QY 95 AMVLAYTKRAARLYTSEYTMNFFVALYLANDMEDEEDYKYLFPWALGDSWRELFP 154
 Db 1202 LVLAKLAKISQYVF 154
 QY 155 LRLQDQ FVAKMTGAVVSEETLVENKFLHAWLRKRGMAHSGAMRYELNEED 216
 Db 1247 VALANELAAKPELSQGRENIAKLYMD 154
 QY 211 FQDQD QTHASQTHKAWTSGVSHNSQSDQLTHVYND 255
 Db 1280 LKAPYELLGTHV---MGV ENLSNHHVTHIR 143
 RESULT 15
 G72724
 probable penicillin acylase; AF04346 Acetopyrum ferriox (Strain K1)
 C:Species: Acetopyrum ferriox
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #taxi_change 20 Aug 1999
 C:Accession: G72724
 K.Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Nakagawa, Y.; Ito, K.; Ito,
 A.; Hara, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper thermophilic cyanobacterium, Acet-
 A:Reference number: A72450; MIM:99410359
 A:Accession: G72724
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-780 <KAW>
 A:Cross-references: DDBJ:AP0000059; NID:q5103911; PDB:BAW7291.1; PDB:11043022; PDB:4
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: AF04346

Query Match 5.2% Score 84.5; DB 2; Length 780;
 Best Local Similarity 22.4% Pred. No. 22;
 Matches 53; Conservative 27; Mismatches 79; Indels 79; Gaps 14;
 QY 65 QAFYRLLENEQTEFLSMNSGLRISDKYLIAMVLAYTKRAARLYTSEYTMNFF 117
 Db 509 ERYITLQERSAEPELDAGMDHQLDVEDLRLKTLVSLMVVASGEGHIVYM 566
 QY 118 FVALYLANDMEDEEDYKYLFPWALGDSWRELFPDFLRDRFPWALGDSWRELFP 177
 Db 567 LALWENQDNNVLYKDFL ALIASW YKAR DFW 568
 QY 178 FVNSKEPTIHWLRKRGMAHSGAMRYELNEED LFTFEG FSLTASGLTHKAWV 242
 Db 599 ESLYSGEENHWFKLE--VAEKALKAVLRGHWVNTPEPTLEGMVNGKVERAVK 652
 QY 243 DSGVSHNSQSDQLTHVYND 255
 Db 653 -----TSETFASNPFAWVYGMHYND 654
 HLLSEAL 684

Search completed: July 3, 2002, 11:09:11
Job time: 4 hr 50c




```

1  APPLICANT: Fischer, Horst
2  APPLICANT: Illek, Beate
3  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC
4  TITLE OF INVENTION: FIBROSIS THERAPY
5  NUMBER OF SEQUENCES: 6
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: SED and HERRY LLP
8  STREET: 6300 Columbia Center, 701 Fifth Avenue
9  CITY: Seattle
10 STATE: Washington
11 COUNTRY: USA
12 ZIP: 98104
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC DOS/MS-DOS
17 SOFTWARE: PatentIn Release #1.0, Version #1.40
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/951,912
20 FILING DATE: 16-Oct-1997
21 CLASSIFICATION: 514
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Makl, David J.
24 REGISTRATION NUMBER: 31,392
25 REFERENCE/DOCKET NUMBER: 200116.403
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: (206) 522-4900
28 TELEFAX: (206) 682-6041
29 INFORMATION FOR SEQ ID NO: 4:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 1479 amino acids
32 TYPE: amino acid
33 TOPOLOGY: linear
34 MOLECULE TYPE: protein
35 US-08-951-912-4

```

```

Query Match          5.38; Score 86; DB 2; Length 1479;
Best Local Similarity 20.7%; Pred. No. 2;
Matches 42; Conservative 47; Mismatches 74; Indels 50; Gaps 8;

QY 9 RASSICGSGVKQVI GKQGHVHAPVVIARRAQIDPERFELSVKQKMKVNIHLNIQPOE 63
DB 220 QASAFQGLGELIVLAIQAGLGKMMKYPQPKUSER--IVLSSEMLNIGSVKAYCW 277
QY 64 KQATYRLLENQIQEFTLSMSGLRISKYLIAVMVLAAYKRAAGLYT
DB 278 FEAMKMIENLRQTE-----LKLTKK---AAVVRFNSSAFFESGFFVWFLSVLPYAL 327
QY 110 ---SEYTMNFEVALTIANIMEDDEHYKYEFPWALGUSWHELFPEFLRDEFW 162
DB 428 IKGILLKIKITITISFCIVLRMVTRO-----FPMVAV QTWYDSLGATNKIQ DEL 375
QY 163 AKMYHRAVSRRCDDVMSKDIPT 185
DB 376 OKQYKTELYNNTTTEVVMENV 398

RESULT 3
US-09-174-077-4
1 Sequence 4, Application US/09174077
2 Patent No. 6329422
3 GENERAL INFORMATION:
4 APPLICANT: Fischer, Horst
5 APPLICANT: Illek, Beate
6 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC FIBROSIS THERAPY
7 FILE REFERENCE: 200116.404C1
8 CURRENT APPLICATION NUMBER: US/09/174,077
9 CURRENT FILING DATE: 1998-10-16
10 EARLIER APPLICATION NUMBER: US 08/951,912
11 EARLIER FILING DATE: 1997-10-16
12 NUMBER OF SEQ ID NOS: 6
13 SOFTWARE: PatentIn Ver. 2.0

```

```

Query Match          5.38; Score 86; DB 4; Length 1479;
Best Local Similarity 20.7%; Pred. No. 2;
Matches 42; Conservative 47; Mismatches 74; Indels 50; Gaps 8;

QY 9 RASSICGSGVKQVI GKQGHVHAPVVIARRAQIDPERFELSVKQKMKVNIHLNIQPOE 63
DB 220 QASAFQGLGELIVLAIQAGLGKMMKYPQPKUSER--IVLSSEMLNIGSVKAYCW 277
QY 64 KQATYRLLENQIQEFTLSMSGLRISKYLIAVMVLAAYKRAAGLYT
DB 278 FEAMKMIENLRQTE-----LKLTKK---AAVVRFNSSAFFESGFFVWFLSVLPYAL 327
QY 110 ---SEYTMNFEVALTIANIMEDDEHYKYEFPWALGUSWHELFPEFLRDEFW 162
DB 328 IKGILLKIKITITISFCIVLRMVTRO-----FPMVAV QTWYDSLGATNKIQ DEL 375
QY 163 AKMYHRAVSRRCDDVMSKDIPT 185
DB 376 OKQYKTELYNNTTTEVVMENV 398

RESULT 4
US-07-637-621-2
1 Sequence 2, Application US/97647621
2 Patent No. 5407796
3 GENERAL INFORMATION:
4 APPLICANT: cutting, daty
5 APPLICANT: antonarakis, stylimos e
6 APPLICANT: kazazian ir, baig h
7 TITLE OF INVENTION: CYSTIC FIBROSIS MUTATION CLUSTER
8 NUMBER OF SEQUENCES: 4
9 CORRESPONDENCE ADDRESS:
10 ADDRESSEE: Panter, Birch, McKie and Beckett
11 STREET: 1001 G Street, N.W.
12 CITY: Washington, D.C.
13 COUNTRY: USA
14 ZIP: 20001
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: PatentIn Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/07/647,621
22 FILING DATE: 19910104
23 CLASSIFICATION: 4C5
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Kanda, Satoh a
26 REGISTRATION NUMBER: 42,141
27 REFERENCE/DOCKET NUMBER: 1107,040D10
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: 202-508 9100
30 TELEFAX: 202-508-9100
31 INFORMATION FOR SEQ ID NO: 2:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 1480 amino acids
34 TYPE: AMINO ACID
35 TOPOLOGY: linear
36 MOLECULE TYPE: protein
37 ORIGINAL SOURCE:
38 ORGANISM: HOMO SAPIENS
39 US-07-637-621-2

Query Match          5.38; Score 86; DB 1; Length 1480;
Best Local Similarity 20.7%; Pred. No. 2;

```

```

Matches 42: Conservative 3 M. matches 74: Indels 50: Gaps 8:
QY 9 KASSICSGVKQVI-----GKCHAAKVAKALDEPEELSVKPKMVPNHNILQPOE 63
DB 220 QASAPCGIGLIVLAIFQCHGMMKKYKQAGKSER-IVTISEMIENISVKAYCW 277
QY 64 RQAFYRLLENEQIFELSMSCSTIRISKYLIAMVLAIFKRAAGLYT----- 109
DB 278 EEMAKMIENLRQTE-----LKLTRK--AAVVRYNSSAFESFEFVVLVLPVAL 327
QY 110 -----SYTITMNFVALYLANIMEEREDYKYLFIFFWALGDSWRLPQFELRLDDPW 162
DB 428 IKGITLRIKFIETTSFIVLPMVITPQ-----EPWAV-QTWYDSLGAINKIQ-DFL 375
QY 163 AKMNPVAVVSPGCDDEVMSKDEPT 185
DB 376 QKQYKTILEYNLITTEVVMNVI 398
QY 163 AKMNPVAVVSPGCDDEVMSKDEPT 185
DB 376 QKQYKTILEYNLITTEVVMNVI 398
RESULT 6
US-08-135-809A-2
Sequence 2, Application US/08135809A
Patent No. 5686677
GENERAL INFORMATION:
APPLICANT: CHENG, SENG H.
APPLICANT: DITULLIO, PAUL
APPLICANT: EBERI, KARL M.
APPLICANT: MEADE, HARRY M.
APPLICANT: SMITH, ALAN E.
TITLE OF INVENTION: DEOXYRIBONUCLEIC ACIDS CONTAINING
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/135,809A
FILING DATE: 13-OCT-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: 104-9,12
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1480 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-135-809A-2
Query Match 5.3%, Score 86, DB 1, length 1480:
Best Local Similarity 20.7%, Pred. No. 2;
Matches 42: Conservative 37: Mismatches 74: Indels 50: Gaps 8:
QY 9 KASSICSGVKQVI-----GKCHAAKVAVVQAPKALGPEELSVKPKMVPNHNILQPOE 63
DB 220 QASAPCGIGLIVLAIFQCHGMMKKYKQAGKSER-IVTISEMIENISVKAYCW 277
QY 64 RQAFYRLLENEQIFELSMSCSTIRISKYLIAMVLAIFKRAAGLYT----- 109
DB 278 EEMAKMIENLRQTE-----LKLTRK--AAVVRYNSSAFESFEFVVLVLPVAL 327
QY 110 -----SYTITMNFVALYLANIMEEREDYKYLFIFFWALGDSWRLPQFELRLDDPW 162
DB 428 IKGITLRIKFIETTSFIVLPMVITPQ-----EPWAV-QTWYDSLGAINKIQ-DFL 375
QY 163 AKMNPVAVVSPGCDDEVMSKDEPT 185
DB 376 QKQYKTILEYNLITTEVVMNVI 398
QY 163 AKMNPVAVVSPGCDDEVMSKDEPT 185
DB 376 QKQYKTILEYNLITTEVVMNVI 398
RESULT 6
US-08-136-742A-2
Sequence 2, Application US/ 818,42A
Patent No. 5670488
GENERAL INFORMATION:
APPLICANT: Gregory, R.J., Applicant, B.J. Couture, L.A., Smith,
APPLICANT: A.E.
TITLE OF INVENTION: GENETICALLY DERIVED CYCLIC PEPTIDES
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: BROMHAUGH, GRAV-S, I WHITE & RAYMOND
STREET: 30 ROCKEFELLER PLAZA
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 07/445,413
FILING DATE: 02 DEC 1992
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Selde, Rochelle K.
REGISTRATION NUMBER: 23,300
REFERENCE/DOCKET NUMBER: A-669 Contingent F.P. 104-9,113
TELEPHONE: (212) 408-2500
TELEFAX: (212) 765-2519
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1480 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-136-742A-2
Query Match 5.3%, Score 86, DB 1, length 1480:
Best Local Similarity 20.7%, Pred. No. 2;
Matches 42: Conservative 37: Mismatches 74: Indels 50: Gaps 8:
QY 9 KASSICSGVKQVI-----GKCHAAKVAVVQAPKALGPEELSVKPKMVPNHNILQPOE 63
DB 220 QASAPCGIGLIVLAIFQCHGMMKKYKQAGKSER-IVTISEMIENISVKAYCW 277
QY 64 RQAFYRLLENEQIFELSMSCSTIRISKYLIAMVLAIFKRAAGLYT----- 109
DB 278 EEMAKMIENLRQTE-----LKLTRK--AAVVRYNSSAFESFEFVVLVLPVAL 327
QY 110 -----SYTITMNFVALYLANIMEEREDYKYLFIFFWALGDSWRLPQFELRLDDPW 162
DB 428 IKGITLRIKFIETTSFIVLPMVITPQ-----EPWAV-QTWYDSLGAINKIQ-DFL 375
QY 163 AKMNPVAVVSPGCDDEVMSKDEPT 185
DB 376 QKQYKTILEYNLITTEVVMNVI 398
QY 163 AKMNPVAVVSPGCDDEVMSKDEPT 185
DB 376 QKQYKTILEYNLITTEVVMNVI 398

```

```

1b 278 EEMAKMIENLRQTE-----LKLTRK--AAVVRYNSSAFESFEFVVLVLPVAL 327
QY 110 -----SYTITMNFVALYLANIMEEREDYKYLFIFFWALGDSWRLPQFELRLDDPW 162
DB 328 IKGITLRIKFIETTSFIVLPMVITPQ-----EPWAV-QTWYDSLGAINKIQ-DFL 375
QY 163 AKMNPVAVVSPGCDDEVMSKDEPT 185
DB 376 QKQYKTILEYNLITTEVVMNVI 398
RESULT 6
US-08-135-809A-2
Sequence 2, Application US/08135809A
Patent No. 5686677
GENERAL INFORMATION:
APPLICANT: CHENG, SENG H.
APPLICANT: DITULLIO, PAUL
APPLICANT: EBERI, KARL M.
APPLICANT: MEADE, HARRY M.
APPLICANT: SMITH, ALAN E.
TITLE OF INVENTION: DEOXYRIBONUCLEIC ACIDS CONTAINING
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/135,809A
FILING DATE: 13-OCT-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: 104-9,12
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1480 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-135-809A-2
Query Match 5.3%, Score 86, DB 1, length 1480:
Best Local Similarity 20.7%, Pred. No. 2;
Matches 42: Conservative 37: Mismatches 74: Indels 50: Gaps 8:
QY 9 KASSICSGVKQVI-----GKCHAAKVAVVQAPKALGPEELSVKPKMVPNHNILQPOE 63
DB 220 QASAPCGIGLIVLAIFQCHGMMKKYKQAGKSER-IVTISEMIENISVKAYCW 277
QY 64 RQAFYRLLENEQIFELSMSCSTIRISKYLIAMVLAIFKRAAGLYT----- 109
DB 278 EEMAKMIENLRQTE-----LKLTRK--AAVVRYNSSAFESFEFVVLVLPVAL 327
QY 110 -----SYTITMNFVALYLANIMEEREDYKYLFIFFWALGDSWRLPQFELRLDDPW 162
DB 428 IKGITLRIKFIETTSFIVLPMVITPQ-----EPWAV-QTWYDSLGAINKIQ-DFL 375
QY 163 AKMNPVAVVSPGCDDEVMSKDEPT 185
DB 376 QKQYKTILEYNLITTEVVMNVI 398
QY 163 AKMNPVAVVSPGCDDEVMSKDEPT 185
DB 376 QKQYKTILEYNLITTEVVMNVI 398

```

QY 163 AKMYKAVVSRCCDEVMSKPT 185
 DB 376 QKQYKTLFYNLTTEVVMENVT 398
 RESULT 7
 US-08-466-886-17
 : Sequence 17, Application US/08456886
 : Patent No. 5776677
 : GENERAL INFORMATION:
 : APPLICANT: Tsui, Lap Chee
 : APPLICANT: Riordan, John R.
 : APPLICANT: Rommens, Johanna M.
 : APPLICANT: Kerem, Bat-Sheva
 : APPLICANT: Collins, Francis S.
 : APPLICANT: Iannuzzi, Michael C.
 : APPLICANT: Drumm, Mitchell L.
 : APPLICANT: Buckwald, Manuel
 : TITLE OF INVENTION: Cystic Fibrosis Gene
 : NUMBER OF SEQUENCES: 43
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX
 : STREET: 1100 New York Avenue, N.W.
 : CITY: Washington
 : STATE: DC
 : COUNTRY: USA
 : ZIP: 20005
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.40
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/466,886
 : FILING DATE: 06-JUN-1995
 : CLASSIFICATION: 4.45
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Goldstein, Jorge A.
 : REGISTRATION NUMBER: 29,021
 : REFERENCE/DOCKET NUMBER: 1329,0010006
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 202-371-2600
 : TELEFAX: 202-371-2540
 : INFORMATION FOR SEQ ID NO: 17:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1480 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08 466-886-17

Query Match 5.38; Score 86; DB 1; Length 1480;
 Best Local Similarity 20.7%; Pred. No. 2;
 Matches 42; Conservative 47; Mismatches 74; Indels 50; Gaps 8;
 QY 9 RASSIGSGVKQVI-----GKGHHAVVVGAKKAGIDPEELLSVKMKVYNTHINLQVDF 63
 DB 220 QASAFGLGLFVLIALFDAGLGMKKRYRQRAKISER IVTISEMITENTQSVKAYVW 277
 QY 64 KQAFYRIENPQDFEFLMSDGLPISDKYLAMVLAVKRAALYLT 109
 DB 278 FRAMKMKIENLRQIE-----LKLTKR---AAYVFNSSAFPSGPFVFLSVLPVAL 327
 QY 110 -----SEYITMFFVVALYLANIMEDREDDYKYLEFPWALDISWRELFPQFLRLRDDFW 162
 DB 328 IKGILIRKIETFTISGVILRMVATRO-----FPWAV QTWYDLSGLINKIQ-DFL 375
 QY 163 AKMYKAVVSRCCDEVMSKPT 185
 DB 376 QKQYKTLFYNLTTEVVMENVT 398

RESULT 8
 US-08-951-912-2
 : Sequence 2, Application US/08951912
 : Patent No. 5972995
 : GENERAL INFORMATION:
 : APPLICANT: Fischer, Horst
 : APPLICANT: Illick, Beate
 : TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC
 : TITLE OF INVENTION: FIBROSIS THERAPY
 : NUMBER OF SEQUENCES: 6
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: SEED and BERRY LLP
 : STREET: 6400 Columbia Center, 701 Fifth Avenue
 : CITY: Seattle
 : STATE: Washington
 : COUNTRY: USA
 : ZIP: 98104
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.40
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: 08/951,912
 : FILING DATE: 16-OCT-1997
 : CLASSIFICATION: 514
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Makl, David J.
 : REGISTRATION NUMBER: 41,492
 : REFERENCE/DOCKET NUMBER: 200116,404
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (206) 622-4900
 : TELEFAX: (206) 682-6041
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1480 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08 951-912-2

Query Match 5.38; Score 86; DB 2; Length 1480;
 Best Local Similarity 20.7%; Pred. No. 2;
 Matches 42; Conservative 47; Mismatches 74; Indels 50; Gaps 8;
 QY 9 RASSIGSGVKQVI-----GKGHHAVVVGAKKAGIDPEELLSVKMKVYNTHINLQVDF 63
 DB 220 QASAFGLGLFVLIALFDAGLGMKKRYRQRAKISER IVTISEMITENTQSVKAYVW 277
 QY 64 KQAFYRIENPQDFEFLMSDGLPISDKYLAMVLAVKRAALYLT 109
 DB 278 FRAMKMKIENLRQIE-----LKLTKR---AAYVFNSSAFPSGPFVFLSVLPVAL 327
 QY 110 -----SEYITMFFVVALYLANIMEDREDDYKYLEFPWALDISWRELFPQFLRLRDDFW 162
 DB 328 IKGILIRKIETFTISGVILRMVATRO-----FPWAV QTWYDLSGLINKIQ-DFL 375
 QY 163 AKMYKAVVSRCCDEVMSKPT 185
 DB 376 QKQYKTLFYNLTTEVVMENVT 398
 RESULT 9
 US-08-951-912-6
 : Sequence 6, Application US/08951912
 : Patent No. 5972995
 : GENERAL INFORMATION:
 : APPLICANT: Fischer, Horst
 : APPLICANT: Illick, Beate
 : TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC
 : TITLE OF INVENTION: FIBROSIS THERAPY
 : NUMBER OF SEQUENCES: 6

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: SEED and BERRY P
3 STREET: 6300 Columbia Center, 7 Fil in Ave. #
4 CITY: Seattle
5 STATE: Washington
6 COUNTRY: USA

7 ZIP: 98104
8 COMPUTER READABLE FORM:
9 MEDIUM TYPE: floppy disk
10 COMPUTER: IBM PC compatible
11 OPERATING SYSTEM: PC DOS/MS DOS
12 SOFTWARE: Patent in Release 4.0 Version #1.33
13 CURRENT APPLICATION DATA:
14 FILING DATE: 16-CT-1997
15 CLASSIFICATION: 614

16 ATTORNEY/AGENT INFORMATION:
17 NAME: Maki, David J.
18 REGISTRATION NUMBER: 31,394
19 REFERENCE/DOCKET NUMBER: 20016 104
20 TELECOMMUNICATION INFORMATION:
21 TELEPHONE: (206) 622-4400
22 TELEFAX: (206) 692-6041

23 INFORMATION FOR SEQ ID NO: 6:
24 SEQUENCE CHARACTERISTICS:
25 LENGTH: 1480 amino acids
26 TYPE: amino acid
27 STRANDEDNESS:
28 TOPOLOGY: linear
29 US-08-951-912-4

Query Match 5.3%; Score 86; DB 2; Length 1480;
Best Local Similarity 20.7%, Pred. No. 2;
Matches 42; Conservative 37; Mismatches 74; Indels 50; Gaps 8;

QY 9 RASSICGSGVKQVI-----CKCHPHARVVGAKKAOIKREELSVKPKMKNVNHINIQPQE 63
DB 220 QASAPQRIQFLIVLALFOANIGPMKKYTPCPAGKRISEP--LVITSEMIENT--SVKAYCW 277
QY 64 RQAFYRLLENQIQEFLSMESCLREISKYLLIAMVLAYFKRAAGLYT----- 109
DB 278 RFAEMKMIENLPQPI-----LKLTRK- AAVRVYVNSSAFFPSGFFVFLSVLYAL 427
QY 110 -----SEYTTMNFVVALYLANDMEDEEDYKYEIFPWALGDSWRELFPQFLRDEFW 162
DB 328 IKGIIILPKIFTTISPCIVIPMAVTPQ-----FPWAV-QTWYDSI-AINKIQ-DFL 375
QY 163 AKMNYRAVVSRRKCCDEVMNSKIDPT 185
DB 376 QKQEKLTLEYNLTTTEVVMENVT 398

RESULT 10
US-08-469-461-4
Sequence 2, Application US/08469461H
Patent No. 5981178
GENERAL INFORMATION:
APPLICANT: Tsui, Lap-See
APPLICANT: Rommins, Johanna M.
APPLICANT: Kerem, Bat-Sheva
TITLE OF INVENTION: Introns and Exons of the Cystic Fibrosis Gene and
FILE REFERENCE: 3477-61, 03/477/1984
CURRENT APPLICATION NUMBER: US/08-469-461H
CURRENT FILING DATE: 1995-06-06
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 2
LENGTH: 1480
TYPE: PRI
ORGANISM: Homo sapiens
US-08-469-461-4

Query Match 5.3%; Score 86; DB 2; Length 1480;
Best Local Similarity 20.7%, Pred. No. 2;
Matches 42; Conservative 37; Mismatches 74; Indels 50; Gaps 8;
QY 9 RASSICGSGVKQVI-----CKCHPHARVVGAKKAOIKREELSVKPKMKNVNHINIQPQE 63
DB 220 QASAPQRIQFLIVLALFOANIGPMKKYTPCPAGKRISEP--LVITSEMIENT--SVKAYCW 277
QY 64 RQAFYRLLENQIQEFLSMESCLREISKYLLIAMVLAYFKRAAGLYT----- 109
DB 278 RFAEMKMIENLPQPI-----LKLTRK- AAVRVYVNSSAFFPSGFFVFLSVLYAL 427
QY 110 -----SEYTTMNFVVALYLANDMEDEEDYKYEIFPWALGDSWRELFPQFLRDEFW 162
DB 328 IKGIIILPKIFTTISPCIVIPMAVTPQ-----FPWAV-QTWYDSI-AINKIQ-DFL 375
QY 163 AKMNYRAVVSRRKCCDEVMNSKIDPT 185
DB 376 QKQEKLTLEYNLTTTEVVMENVT 398

RESULT 11
US-08-469-461-4
Sequence 4, Application US/08469461H
Patent No. 5981178
GENERAL INFORMATION:
APPLICANT: Tsui, Lap-See
APPLICANT: Rommins, Johanna M.
APPLICANT: Kerem, Bat-Sheva
TITLE OF INVENTION: Introns and Exons of the Cystic Fibrosis Gene and
FILE REFERENCE: 3477-61, 03/477/1984
CURRENT APPLICATION NUMBER: US/08-469-461H
CURRENT FILING DATE: 1995-06-06
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 4
LENGTH: 1480
TYPE: PRI
ORGANISM: Homo sapiens
US-08-469-461-4

Query Match 5.3%; Score 86; DB 2; Length 1480;
Best Local Similarity 20.7%, Pred. No. 2;
Matches 42; Conservative 37; Mismatches 74; Indels 50; Gaps 8;
QY 9 RASSICGSGVKQVI-----CKCHPHARVVGAKKAOIKREELSVKPKMKNVNHINIQPQE 63
DB 220 QASAPQRIQFLIVLALFOANIGPMKKYTPCPAGKRISEP--LVITSEMIENT--SVKAYCW 277
QY 64 RQAFYRLLENQIQEFLSMESCLREISKYLLIAMVLAYFKRAAGLYT----- 109
DB 278 RFAEMKMIENLPQPI-----LKLTRK- AAVRVYVNSSAFFPSGFFVFLSVLYAL 427
QY 110 -----SEYTTMNFVVALYLANDMEDEEDYKYEIFPWALGDSWRELFPQFLRDEFW 162
DB 328 IKGIIILPKIFTTISPCIVIPMAVTPQ-----FPWAV-QTWYDSI-AINKIQ-DFL 375
QY 163 AKMNYRAVVSRRKCCDEVMNSKIDPT 185
DB 376 QKQEKLTLEYNLTTTEVVMENVT 398

RESULT 12
US-08-469-461-4
Sequence 2, Application US/084694605
Patent No. 5981714
GENERAL INFORMATION:
APPLICANT: Cheng, Song H., Marshall, John, Gregory, Richard J.
APPLICANT: and Rafter, Patrick W.

```

: TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR CYSTIC FIBROSIS
: TITLE OF INVENTION: TRANSMEMBRANE CONDUCTANCE REGULATOR AND USES
: TITLE OF INVENTION: THEREFOR
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & CUCKFIELD
: STREET: 60 STATE STREET, SUITE 510
: CITY: BOSTON
: STATE: MASSACHUSETTS
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/691,605
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/114,950
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hanley, Elizabeth A.
: REGISTRATION NUMBER: 33,505
: REFERENCE/DOCKET NUMBER: NZ1-029
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 227-5941
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1480 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-691-605-2

```

```

Query Match: 5.48; Score 86; DB 2; Length 1480;
Best Local Similarity: 20.7%; Pred. No. 2;
Matches 42; Conservative 47; Mismatches 74; Indels 50; Gaps 8;

QY 9 RASSICGSGVKQVI-----CKGHPHARVVCARKAKQIPERFEELSVKPKVNRNHLNLPQE 63
DB 220 CASAFCCGRLVIAAQA:GCHMMKYNFCAGKISER LVTSSEMENLQSVKAYCW 277
QY 64 RQAFYRLLENEQTEFLSMDSCLRSISKYLLAMVLAYFKRAAGLYI----- 109
DB 278 FEAMKMIENLRQTE-----LKLTRK---AAYVYFNSSAFFSGFVVFLSVLPYAL 327
QY 110 -----SEYITNNFVVALYLANIMEDEEDYKYIEFWALGDSWPKFPGFPIRPPFW 162
DB 328 IKGILLKRIPTTISFVILRMVAVTQ-----FWAVQTWYDSLGAIKNTQ-DFL 475
QY 163 AKMYRAVVSRRCCDVMMSKPT 185
DB 376 QKQYKILYNYLITTEVVMENVI 398

```

```

RESULT 13
US-08 455-552A-14
: Sequence 14, Application US/08455552A
: Patent No. 5990279
: GENERAL INFORMATION:
: APPLICANT: Carter, Barrie J.
: APPLICANT: Flotte, Terence
: APPLICANT: Afione, Sandra
: APPLICANT: Solow, Ricki
: TITLE OF INVENTION: MODIFIED ADENO ASSOCIATED VIRUS VECTOR
: TITLE OF INVENTION: CAPABLE OF EXPRESSION FROM A NOVEL PROMOTER
: NUMBER OF SEQUENCES: 14

```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NEEDLE & ROSENBERG
: STREET: 127 Peachtree Street, Suite 1200
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.2A
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: 95/09,455,552A
: FILING DATE: 31 May 1995
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Pettyjohn, David G.
: REGISTRATION NUMBER: 65,438
: REFERENCE/DOCKET NUMBER: 20094-0192
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (404) 688 0770
: TELEFAX: (404) 688 9880
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1480 amino acids
: TYPE: amino acid
: STRANDNESS: not relevant
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-455-552A-14

```

```

Query Match: 5.48; Score 86; DB 2; Length 1480;
Best Local Similarity: 20.7%; Pred. No. 2;
Matches 42; Conservative 47; Mismatches 74; Indels 50; Gaps 8;

QY 9 RASSICGSGVKQVI-----CKGHPHARVVCARKAKQIPERFEELSVKPKVNRNHLNLPQE 63
DB 220 CASAFCCGRLVIAAQA:GCHMMKYNFCAGKISER LVTSSEMENLQSVKAYCW 277
QY 64 RQAFYRLLENEQTEFLSMDSCLRSISKYLLAMVLAYFKRAAGLYI----- 109
DB 278 FEAMKMIENLRQTE-----LKLTRK---AAYVYFNSSAFFSGFVVFLSVLPYAL 327
QY 110 -----SEYITNNFVVALYLANIMEDEEDYKYIEFWALGDSWPKFPGFPIRPPFW 162
DB 328 IKGILLKRIPTTISFVILRMVAVTQ-----FWAVQTWYDSLGAIKNTQ-DFL 475
QY 163 AKMYRAVVSRRCCDVMMSKPT 185
DB 376 QKQYKILYNYLITTEVVMENVI 398

```

```

RESULT 14
US-07-890-609-2
: Sequence 2, Application US/07890609
: Patent No. 6001588
: GENERAL INFORMATION:
: APPLICANT: Tsui, Lap-Chew
: APPLICANT: Romming, Johanna M.
: APPLICANT: Keren, Bat Sheva
: TITLE OF INVENTION: Infrons and Exons of the Cystic Fibrosis Gene and
: TITLE OF INVENTION: Mutations at Various Positions of the Gene
: FILE REFERENCE: 477/61, 034477/139840
: CURRENT APPLICATION NUMBER: US/07/890,609
: CURRENT FILING DATE: 1992 07 13
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO: 2
: LENGTH: 1480
: TYPE: PRT
: ORGANISM: Homo sapiens

```

$$US - (17 - 889) = 6119 \text{ } \mathbf{2}$$

```

Query Match: 5, 8; Score 65; H: 3; Length 1480;
Hcst local similarity 20, 37; P: 0.2
Matches 42; Conservative 74; Idels 50; Gaps 8;

QY 5 KASSYGSNGVVI-----KHFAPVAPFAA*FEFEESVKKPVNTILNQPE 63
DB 229 SASA*CGHGLPVLVIAFGALGPMNKYP*PAK*SEP--VILSMENIQSVKVCW 277

QY 64 PAAETVPLENQIGFFISMSQTF*SKY*AMVAVYFAAGVYT----- 109
DB 278 EFANEKMIENLRQTF--*LKG*KKAAVAVYFHSAPFFSFFVFLSLVPYAL 327

QY 112 -SEYITMNIENVALYANIN*FEFQ*SYKE*FWALGHSNKEFLPQPI*LRDHSFW 162
DB 328 *K*::ILRKIFITISGIVLMAVTE*...FFWAV*QTWYDLSGAINKIQ*DFL 375

QY 163 ARMYRAVVSRCGLEVMSK*FT*AG 163
DB 376 QKQVKTLEYNI*TF*EVMM*NV* 368

```

```

RESULT 25
US-07-890-604-4
? Sequence 4, Application US/07890 4C
? Patent No. 6001588
? GENERAL INFORMATION:
? APPLICANT: Tsui, Lap-see
? APPLICANT: Kommins, J. Hanna M.
? APPLICANT: Kereem, Bat Sheva
? TITLE OF INVENTION: Tifrons and
? TITLE OF INVENTION: Mutations at Val-
? FILE REFERENCE: 3477-61, 0-3477, 1983
? CURRENT APPLICATION NUMBER: US/07890 604C
? CURRENT FILING DATE: 1992-07-13
? NUMBER OF SEQ ID NOS: 33
? SOFTWARE: PatentIn Ver. 2.1.
? SEQ ID No 4
? LENGTH: 1480
? TYPE: PRT
? ORGANISM: Homo sapiens
US-07-890-604-4

```

[illegible]

Search completed: July 31, 2012, 1:08:
Job time: 4422 sec



DR WPI: 2001-639362/73
 DR N-PSDB: AAS78572
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 XX (MYSE-) HYSEQ INC.
 PS Claim 20, SEQ ID NO 44744; 10pp; English
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. Ab00010 AAG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/publshed_pat_sequences
 XX
 SQ Sequence 391 AA:
 Query Match: 16.7% Score 273; DE 22; Length 391;
 Best local Similarity 36.3%; Pred. No. 2, 16 20;
 Matches 75; Conservative 32; Mismatches 72; Indels 28; Gaps 6;
 QY 35 PKAIPEREEL-----SVKPKMVPNTHLNQDPQFQAFYPLENQLIEFLSMGSCIP 87
 DB 159 rklapepewkw amllstikkkkkrtrssslpdeedafllledpfrlfrfakll 218
 QY 88 ISDKYLIAWVLAYPEKRAAGLYISPTVMNFVALYANDMEDEEDYKYIFPWALGDSW 147
 DB 219 vskyllam:laylar qqlfswqyqlhflalylandeededkq:itflfak 276
 QY 148 RELFPQFLRLKDDFWAKMNYAVVS-----RNEELFFPRG 215
 DB 277 rslptlrrkrifqlyrsmprrknrshlplvrkrifqtrremprknrshlplvifqr 336
 QY 194 PMHH--SGAMRGYL---RNEELFFPRG 215
 DB 437 rrlfcsmscravwspeeleentprq 363
 PS017 4
 AB016876
 ID AB016876 standard: protein: 367 AA
 XX
 AC AB016876
 XX
 DT 18 FEB 2002 (first entry)
 XX
 DE Novel human diagnostic protein: #167-7.
 XX
 KW Human: chromosome mapping; gene mapping; gene therapy; forensics;
 KW food supplement; medical mapping; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN W020175667-A2.
 XX

DR 11-OCF-2001.
 XX
 XX 30-MAR-2001; 2001WO-US086631.
 XX
 XX 21-MAR-2000; 2000US-0540317.
 XX
 XX 23-APR-2000; 2000US-0949167.
 XX
 XX (MYSE-) HYSEQ INC.
 XX
 XX Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI: 2001-639362/73.
 XX N-PSDB: AAS81063.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity
 XX
 XX Claim 20; SEQ ID NO 47235; 10pp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
 XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 XX and gene mapping, and in recombinant production of (II). The
 XX polynucleotides are also used in diagnostics as expressed sequence tags
 XX for identifying expressed genes. (I) is useful in gene therapy techniques
 XX to restore normal activity of (II) or to treat disease states involving
 XX (II). (II) is useful for generating antibodies against it, detecting or
 XX quantitating a polypeptide in tissue, as molecular weight markers and as
 XX a food supplement. (II) and its binding partners are useful in medical
 XX imaging of sites expressing (II). (I) and (II) are useful for treating
 XX disorders involving aberrant protein expression or biological activity.
 XX The polypeptide and polynucleotide sequences have applications in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits to assess biodiversity
 XX and to produce other types of data and products dependent on DNA and
 XX amino acid sequences. Ab00010 AAG30377 represent novel human
 XX diagnostic amino acid sequences of the invention.
 XX Note: The sequence data for this patent did not appear in the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/publshed_pat_sequences.
 XX
 XX Sequence 391 AA:
 Query Match: 16.7% Score 273; DE 22; Length 391;
 Best local Similarity 36.3%; Pred. No. 2, 16 20;
 Matches 75; Conservative 32; Mismatches 72; Indels 28; Gaps 6;
 QY 35 PKAIPEREEL-----SVKPKMVPNTHLNQDPQFQAFYPLENQLIEFLSMGSCIP 87
 DB 159 rklapepewkw amllstikkkkkrtrssslpdeedafllledpfrlfrfakll 218
 QY 88 ISDKYLIAWVLAYPEKRAAGLYISPTVMNFVALYANDMEDEEDYKYIFPWALGDSW 147
 DB 219 vskyllam:laylar qqlfswqyqlhflalylandeededkq:itflfak 276
 QY 148 RELFPQFLRLKDDFWAKMNYAVVS-----RNEELFFPRG 215
 DB 277 rslptlrrkrifqlyrsmprrknrshlplvrkrifqtrremprknrshlplvifqr 336
 QY 194 PMHH--SGAMRGYL---RNEELFFPRG 215
 DB 437 rrlfcsmscravwspeeleentprq 363
 PS017 5
 AB000498
 ID AB000498 standard: protein: 367 AA.
 XX
 AC AB000498
 XX

RESULTS	6
ABG02841	
ID	ABG02841 standard; Protein; 467 AA
XX	
AC	ABG02841;
XX	
D1	13-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #2842.
XX	
KW	human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	W0200175067 A2.
XX	
PP	11-OCT 2001.
XX	
PR	30-MAR 2001; 2001WO 0508641.
XX	
PR	31-MAR 2000; 2000US-0540217.
XX	
PR	23-AUG 2000; 2000US 0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
DR	WP1: 2001-6,9546,2/74.
XX	
DR	N-PSDB; AAS67028.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.
XX	
PS	Claim 20; SEQ ID No 33206; 103pp; English.
XX	
CCC	The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligonucleotides, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detection or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AB0200010 AAC60377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_pat_sequences .
XX	
SSQ	Sequence. 467 AA;

[illegible]

CC at ftp.wipo.int/pub/published_pet_sequences.

XX Sequence 434 AA:

Query Match
Best Local Similarity 16.6% Score 271; DB 22; Length 444;
Matches 71; Conservative 25; Mismatches 47; Gaps 4.

QY 35 RKAAIPEPEFT-----SVKPKMVPNTHLNLOPQRCATVETLENFCIGEFISMESCLR 87
DB 227 kvlaapeetwacmteqikmkikrrvsivlpethcafnllcpvkrflawdkdlr 286

QY 88 ISDKYLLAMVAYKRAAGLYTSYTTMNTFVALYLANDEPDEYKYELFPWALGDS- 146
DB 287 vskkyllamvaylar-aaipswyqrlhllalylandmcedab-qkqnll-ylyqktr 345

QY 147 -----WRELEFPQ-----FLRDRDFWAKMNYKAVVSRCQIDE 178
DB 346 srplvrnrtrqlerehnparkursqialfkrlrfesmsqrawvsteetee 400

RESULT 9
ARG16875
ID ARG16875 Standard; Protein; 434 AA.

XX ARG16875;
XX
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #16866.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.

XX W0200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08641.
XX
XX 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSF-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPT: 2001-639362/73.
XX DR N-PSDB; AAS81062.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.

XX Claim 20; SEQ ID No 47244; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX the polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ARG0010 ARG0077 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: the sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from Wipo
CC at ftp.wipo.int/pub/published_pet_sequences.

XX Sequence 434 AA:

Query Match
Best Local Similarity 16.6% Score 271; DB 22; Length 444;
Matches 71; Conservative 25; Mismatches 47; Gaps 4.

QY 35 RKAAIPEPEEL-----SVKPKMVPNTHLNLOPQRCATVETLENFCIGEFISMESCLR 87
DB 227 kvlaapeetwacmteqikmkikrrvsivlpethcafnllcpvkrflawdkdlr 286

QY 88 ISDKYLLAMVAYKRAAGLYTSYTTMNTFVALYLANDEPDEYKYELFPWALGDS- 146
DB 287 vskkyllamvaylar-aaipswyqrlhllalylandmcedab-qkqnll-ylyqktr 345

QY 147 -----WRELEFPQ-----FLRDRDFWAKMNYKAVVSRCQIDE 178
DB 346 srplvrnrtrqlerehnparkursqialfkrlrfesmsqrawvsteetee 400

RESULT 10
ARG16879
ID ARG16879 Standard; Protein; 411 AA.

XX ARG16879;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #16870.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.

XX W0200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08641.
XX
XX 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.

XX (HYSF-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPT: 2001-639362/73.
XX DR N-PSDB; AAS81066.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.

XX Claim 20; SEQ ID No 47244; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (i). (i) is useful for treating disease states involving
 CC (ii). (ii) is useful for detecting and identifying antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as well as for weight markers and as
 CC a food supplement. (iii) and (iv) are useful in medical
 CC imaging of sites expressing (ii). (v) and (vi) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC the polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AB00010-AB00037 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic form directly from WIPO
 CC at http://wipo.int/pub/published_pat_sequences.

XX Sequence 311 AA:

Query Match: 15.7% Score 248.5; DB 22; Length 311.
 Best Local Similarity 44.7%; Pred. No. 9.8e-18;
 Matches 67; Conservative 21; Mismatches 45; Indels 17; Gaps 6.

QY 35 RKAOIPPEREL-----SVKPKWVNTHNLQPPQKQAFYRLLENQIQEFSMSQSLR 87
 DB 54 RKVLAPPEPEELWVAEMLCGLMKIKRRVSLVPEHBEALNRLLEDVVKRFLAWKDIR 113
 QY 88 ISDKYLIAWLVAYFKRAAGLYTSEYITMNFVALYLANDMEDEEDYKYEIFPWALGDSW 147
 DB 114 VSDKYLIAMVAYFST-AGFSWYQYGRHFFLYLANDMCCDDSKQNFIFYRKAR 171
 QY 148 RHIFQFLRLKDDPW-----AMNPKAVVVSF 173
 DB 172 RSLPILIRK---WQLQRAMPRARKNR 199

RESULT 11

AB000505
 ID ARG000505 standard; Protein: 417 AA

XX AC AB000505

XX DL 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #496

XX KW Human: chromosome mapping, gene mapping, gene therapy, forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PE 30-MAR-2001; 2001WO-0508631.

XX PP 31-MAR-2000; 2000US-0540217.

XX PP 23-AUG-2000; 2000US-0540117.

XX PA (ATSEQ) ATSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DP WPT: 0001-639362/7?

XX DR N-USDB; AAS64692.

XX PT New isolated polynucleotide and encoded polypeptides useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity

XX PT Claim 20; SEQ ID NO 30864 10 pp. require

XX the invention relates to isolated polynucleotide (i) and
 XX polypeptide (ii) sequences. (i) is useful in hybridisation probes,
 XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 XX and gene mapping, and in recombinant production of (ii). The
 XX polynucleotides are also used in diagnostics as expressed sequence tags
 XX for identifying expressed genes. (i) is useful in gene therapy techniques
 XX to restore normal activity of (ii) or to treat disease states involving
 XX (ii). (ii) is useful for generating antibodies against it, detecting or
 XX quantitating a polypeptide in tissue, as molecular weight markers and as
 XX a food supplement. (iii) and (iv) are useful in medical
 XX imaging of sites expressing (ii). (v) and (vi) are useful for treating
 XX disorders involving aberrant protein expression or biological activity
 XX the polypeptide and polynucleotide sequences have applications in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits to assess biodiversity
 XX and to produce other types of data and products dependent on DNA and
 XX amino acid sequences. AB00010-AB00037 represent novel human
 XX diagnostic amino acid sequences of the invention.
 XX Note: The sequence data for this patent did not appear in the printed
 XX specification, but was obtained in electronic form directly from WIPO
 XX at http://wipo.int/pub/published_pat_sequences.

XX Sequence 417 AA:

Query Match: 15.2% Score 248.5; DB 22; Length 417;
 Best Local Similarity 29.3%; Pred. No. 9.8e-18;
 Matches 76; Conservative 32; Mismatches 78; Indels 73; Gaps 7;

QY 35 RKAOIPPEREL-----SVKPKWVNTHNLQPPQKQAFYRLLENQIQEFSMSQSLR 87
 DB 160 RKVLAPPEPEELWVAEMLCGLMKIKRRVSLVPEHBEALNRLLEDVVKRFLAWKDIR 219
 QY 88 ISDKYLIAWLVAYFKRAAGLYTSEYITMNFVALYLANDMEDEEDYKYEIFP----- 140
 DB 220 VSDKYLIAMVAYFST-AGFSWYQYGRHFFLYLANDMCCDDSKQNFIFYRKAR 278
 QY 141 -----WALGDSW-----RELFPQFLRLR 158
 DB 279 RSLPILIRKRWQLQRAMPRARKNRIPILIRKRLQYRSTNPRARKNRSLPILIRK 338
 QY 159 RQFWANMNYPAVVSPECCDEVMKSNVTHW-----AWLPQRMHHSCHAPGFLPVEDDF 211
 DB 339 RQYRSMNRARKNRSAI VIFKRRHFICSMSCRAWVSPEELEEATPRQVDIQEL 397
 QY 212 RPEFSLIFASTLCHKAG 230
 DB 398 YSNANGR-----HQAG 408

RESULT 12

AB002847
 ID AB002847 standard; Protein: 417 AA.

XX AC AB002847

XX DL 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #2838

XX KW Human, chromosome mapping, gene mapping, gene therapy, forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001

XX PP 30-MAR-2001; 2001WO-0508631.

XX PP 31-MAR-2000; 2000US-0540217.

```

PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX N-PSDB; AAS67034.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX Claim 20; SEQ ID No 33206; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG3037 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 417 AA;

Query Match 15.2%; Score 248.5; DB 23; Length 417;
Best Local Similarity 29.3%; Pred. No. 9.8e-18;
Matches 76; Conservative 32; Mismatches 78; Indels 73; Gaps 7;

QY 35 RKAQIDPEREL-----SVKPKWVNTHTLNLOPQERQAFYRLLENEQTEFLSMDSCLR 87
Db 160 rkvlapepeewwamqlgymkkrkrrvstypethcafnrlledpqlk:fiawdkdlr 219
QY 88 ISKYLILAMVLAFFKKAAGLYTSRYTTMNFVVALYLANDMEEDDEEDYKYVEFP----- 140
Db 226 vskdyllamvlaytsr atfswayqrinfllalylandme-ahbskshqthllylkr 478
QY 141 -----WALGDSW-----RMFPQFLER 158
Db 279 srpllrkrwrlqghsmprarknrslpllrkrfqlyrstprarknrslpllrkr 338
QY 159 DDFWAKMNYRAVVSRRCCDEVMSSKDTIHW-----AWLRDRFMHHSAMRGYLNEDQF 211
Db 339 fqlrysmmrarknrsl-qvlfqkrffhffcsmscravswspsctecntqprgdvdfgql 397
QY 212 FDRGPGHTPASCTLCHKAG 230
Db 398 ysbauqr-----hqaq 408

RESULT 13
ID ABG21399
XX ABG21399 standard; Protein: 440 AA.
AC
XX ABG21399;
XX
DT 18-FEB-2002 (first entry)

```

```

XX DE
XX Novel human diagnostic protein #21399.
XX Human: chromosome mapping; gene mapping; gene therapy; forensics;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX GS
XX Homo sapiens.
XX PN WO200175067 A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08641.
XX PR 31-MAR-2000; 2000US-0640217.
XX PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX N-PSDB; AAS85586.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX Claim 20; SEQ ID No 51758; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG3037 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 440 AA;

Query Match 15.2%; Score 248.5; DB 23; Length 440;
Best Local Similarity 29.3%; Pred. No. 1.1e-17;
Matches 76; Conservative 32; Mismatches 78; Indels 73; Gaps 7;

QY 35 RKAQIDPEREL-----SVKPKWVNTHTLNLOPQERQAFYRLLENEQTEFLSMDSCLR 87
Db 160 rkvlapepeewwamqlgymkkrkrrvstypethcafnrlledpqlk:fiawdkdlr 219
QY 88 ISKYLILAMVLAFFKKAAGLYTSRYTTMNFVVALYLANDMEEDDEEDYKYVEFP----- 140
Db 226 vskdyllamvlaytsr atfswayqrinfllalylandme-ahbskshqthllylkr 478
QY 141 -----WALGDSW-----RMFPQFLER 158
Db 302 srpllrkrwrlqghsmprarknrslpllrkrfqlyrstprarknrslpllrkr 361
QY 159 DDFWAKMNYRAVVSRRCCDEVMSSKDTIHW-----AWLRDRFMHHSAMRGYLNEDQF 211

```


Db	462	tlvysmmsrknrsqtlvrlqr::thfl smsr twvspeelent aprqvdldqgel	420						
QY		: : : : : I I I I I : : : : : I I I I I : : : : :							
QY	212	FREGPLTPASCTLCBKAG 230							
QY		: : : : : I I I I I : : : : : I I I I I : : : : :							
Db	429	yvnangr-----hqag 431							
PS	14								
ARG14386									
XX	XX	ABGI4386 standard; Protein: 410 AA							
XX	XX	ABGI4386;							
DT	18-FEB-2002 (first entry)								
XX	XX	Novel human diagnostic protein #14377.							
XX	XX	Human; chromosome mapping, gene mapping; gene therapy; forensic;							
XX	XX	food supplement; medical imaging; diagnostic; genetic disorder.							
XX	XX	Homo sapiens.							
OS	XX								
PN	XX	WC200175067-A2.							
XX	XX	11-OCT-2001.							
PD	XX	30-MAR-2001; 2001WO-US08631.							
PF	XX	31-MAR-2000; 2000US-0540217.							
PR	XX	23-AUG-2000; 2000US-0649157.							
XX	XX	(HYSEQ-) HYSEQ INC.							
XX	XX	Dmanac RT, Liu C, Tang YT;							
XX	XX	WIPI: 2001-639362/73.							
DR	XX	N-PADB; AAS78573.							
XX	XX	New isolated polynucleotide and encoded polypeptides, useful in							
PPT	XX	diagnostics, forensics, gene mapping; identification of mutations							
PPT	XX	responsible for genetic disorders or other traits and to assess							
PT	XX	biodiversity .							
PS	XX	Claim 20; SEQ ID No 44745; 103pp; English.							
XX	XX	The invention relates to isolated polynucleotide (I) and							
XX	XX	polypeptide (II) sequences. (I) is useful as hybridisation probes,							
CC	XX	polymerase chain reaction (PCR) primers, oligomers, and for chromsome							
CC	XX	and gene mapping, and in recombinant production of (II). The							
CC	XX	polynucleotides are also used in diagnostics as expressed sequence tags							
CC	XX	for identifying expressed genes. (I) is useful in gene therapy techniques							
CC	XX	to restore normal activity, or (II) or to treat disease states involving							
CC	XX	(II). (II) is useful for generating antibodies against it, detecting or							
CC	XX	quantitating a polypeptide in tissue, as molecular weight markers, and as							
CC	XX	a food supplement. (III) and its binding partners are useful in medical							
CC	XX	imaging of sites expressing (II); (I) and (II) are useful for treating							
CC	XX	disorders involving aberrant protein expression or biological activity.							
CC	XX	The polypeptide and polynucleotide sequences have applications in							
CC	XX	diagnostics, forensics, gene mapping; identification of mutations							
CC	XX	responsible for genetic disorders or other traits to assess biodiversity							
CC	XX	and to produce other types of data and facts dependent on DNA and							
CC	XX	diagnostic amino acid sequences of the inventions.							
CC	XX	Note: The sequence data for this patent did not appear in the printed							
CC	XX	specification, but was obtained directly from WIPO							
CC	XX	at fip.wipo.int/poly/publicated.htm#pages							
XX	XX	Sequence 448 AA:							
SSQ									
Query Match:	15/20	Score 148.51	DB 22	Length 448.					
Best Local Similarity	20/20	Prod. 5.0	100-100						

CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG00477 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pat_sequences.

XX
 SQ Sequence 385 AA:

```

Query Match          9.58; Score 154.5; DB 22; Length 385;
Best Local Similarity 29.94; Pred. No. 1e-07;
Matches 55; Conservative 21; Mismatches 61; Indels 47; Gaps 5;

QY 35 KKAQIPEREEL--- --SVKPKWVRNTHINIQPQRCAFYELLENFQGFELSMDSCLR 87
   || || || | | | | | | | | | | | | | | | | | | | | | | | |
Db 164 rkvlapqeeewvacmleqkkmkikrrrslylpchkaafurllqpvrkrflawdkqir 222
      :|||
QY 88 ISDRY -----LIAMVLAYEKRAAGLYTSEYTTMNFVAL-- -YLANIMEELED 133
   :||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 223 vsdkipseptlqaspktilpasricrpsnppsrnfpmstvrpulsylandmceddd 282
      :||| | | | | | | | | | | | | | | | | | | | | | | | |
QY 134 YKYELFWALGDSWRKELEP-----GFLRLRDRHWAKMNYR 168
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 283 pkqityilydkt-rsripllrkrrqleromprarkarsqvllqkrlqltesmscr 341
      | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 169 AVVS 172
   | |
Db 342 awvs 345

```

Search completed: July 31, 2002, 17:07:23
 Job time: 5310 sec


```

Db 121 GATCTCCGGGACGACTTCTGGGTAAATGAACATACGAGCACTGTTATGAGAAAGTG 180
QY 738 etqtdaqtatqatgcacaagaatcccaactcattgaacttaactgaagaatcaccacat 797
Db 181 GTCGATGAGGTAACGTCGAAAGATCCGACACATATGGGCCCTGGGTGACAGATGCCGCAAT 240
QY 798 gcatcacagaacagagatqatqatqatqatqatqatqatqatqatqatqatqatqatqat 857
Db 241 CCAATCACAGCGGAGGCAATGCGTGGGTACCTTACGATGAGGAGGAGGAGGAGGAGGAGG 300
QY 858 tccacacacttaccacacacacacacacacacacacacacacacacacacacacacacacac 917
Db 301 TCGAGGCTTACAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
QY 918 tgggtctccacacacacacacacacacacacacacacacacacacacacacacacacac 977
Db 361 TGGGCTCTCCACACAAACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
QY 978 aaatqatqatqatqatqatqatqatqatqatqatqatqatqatqatqatqatqatqatq 1012
Db 421 GAGTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 455

RESIDUE 15
LOCUS B1444496 534 bp mRNA linear EST 21-AUG-2001
DEFINITION dec25e07.y4 Wellcome CRC pRNA oocyte Xenopus laevis cDNA clone
IMAGE: 3472884 5' similar to IR:Q9VGL1 Q9VGL1 SPEEDY PROTEIN. ;
mRNA sequence.
ACCESSION B1444496
VERSION B1444496.1 GI:15269203
KEYWORDS EST
SOURCE African clawed frog.
ORGANISM
  Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;
  Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
  Xenopodidae; Xenopus.
  1 (bases 1 to 534)
REFERENCE
  Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
  Martin,I., Wyke,T., Underwood,K., Theising,B., Powers,Y., Person,
  B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
  Waterston,R. and Wilson,R.
  WashU Xenopus EST project, 1999
  Unpublished (1999)
  Contact: Sandy Clifton, Ph.D.
  WashU Xenopus EST project, 1999
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, Mo 63108, USA
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est.watson.wustl.edu
  Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.H.
  Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
  University Genome Sequencing Center
  clone distribution: Xenopus clones from this library are available
  through the I.M.A.G.E. Consortium/LLNL at: info:image.llnl.gov
  High quality sequence stop: 429.
  Location/Qualifiers
    1..534
      /organism="Xenopus laevis"
      /db_xref="taxon:8345"
      /clone="IMAGE:3472884"
      /clone_lib="Wellcome CRC pRNA oocyte"
      /rna_seq_type="oocyte"
      /lab_host="DH10B (phage-resistant)"
      /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
      were oligo-dT primed and directionally cloned. Sequencing
      according to Newkooop and Faber. Library was constructed
      by N. Garrett, P. LeMaire, A.M. Zorn, and J.H. Gurdon
      (Wellcome/CRC Institute)."
BASE COUNT 137 a 144 c 129 g 134 t 1 others
ORIGIN

```

Query Match: 27.2%; Score: 4.06; E: 1.0e-10; Length: 534;
 Best Local Similarity: 90.8%; Posit. No.: 126; Neg.
 Matches: 476; Conservative: 0; Mismatches: 45; Indels: 4; Gaps: 1.

```

QY 131 TTTGGGCTGTTGGTGGCTAGGGGCTGAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 190
Db 14 TTTGGGCTGTTGGTGGCTAGGGGCTGAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 190
QY 191 ggggagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 250
Db 74 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 250
QY 251 TTTGGGCTGTTGGTGGCTAGGGGCTGAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 310
Db 134 TTTGGGCTGTTGGTGGCTAGGGGCTGAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 310
QY 311 ggggagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 370
Db 194 GAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 370
QY 371 ggggagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 430
Db 254 GATTTAAATGATTTAAATGATTTAAATGATTTAAATGATTTAAATGATTTAAATGATTTAA 430
QY 431 atgagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 490
Db 314 ATGAGCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 490
QY 491 tgggagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 550
Db 374 TATTTAAATGATTTAAATGATTTAAATGATTTAAATGATTTAAATGATTTAAATGATTTAA 550
QY 551 ggggagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 610
Db 431 CAATGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 610
QY 611 actataatataatataatataatataatataatataatataatataatataatataatataat 670
Db 451 ACTATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 670

```

Search completed: July 31, 2002, 16:11:00
 Job time: 2482 sec

Query Match
 Best Local Similarity 2.4%; Score 37.2; DB 1; Length 7218;
 Matches 6; Conservative 208; Mismatches 156; Indels 0; Gaps 0;

QY 843 ctttttcccgagggtccagaccttaccagccagccagctgtacacctttccataaaagagg 902
 Db 1075 YY 1134
 QY 903 tgtctgttactctgtgggtctccacacacacactctctctccacacacagagattt 962
 Db 1135 YY 1194
 QY 964 tcaclacaccaatagggaatcccaagagactctccatctacaccccttaccatgttct 1022
 Db 1195 YY 1254
 QY 1024 quatcccgagtgactgaacttaccacactctccagagccacttatttaccacac 1082
 Db 1255 YY 1314
 QY 1083 agatggagcggcgtggatggacacacacttcttaccagagcttctcccgagctttt 1142
 Db 1315 YY 1374
 QY 1143 attctctctaccagagctcaagacttattctctctctcaagacttctcaagacttc 1202
 Db 1375 YY 1434
 QY 1203 agacttaatt 1212
 Db 1435 GTACCAATT 1444

RESULT 2

US-08-998-416-224
 : Sequence 224, Application US/08998416
 : Patent No. 6239264
 : GENERAL INFORMATION:
 : APPLICANT: Philippsen, Peter
 : APPLICANT: Pohlmann, Rainer
 : APPLICANT: Stelner, Sabine
 : APPLICANT: Mohr, Christine
 : APPLICANT: Wendland, Jürgen
 : APPLICANT: Knechtle, Philipp
 : APPLICANT: Rebschuh, Corinne
 : TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII
 : TITLE OF INVENTION: AND USES THEREOF
 : NUMBER OF SEQUENCES: 1152
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: No. 6239264artis Corporation
 : STREET: 3054 Cornwallis Road
 : CITY: Research Triangle Park
 : STATE: No. 6239264th Carolina
 : COUNTRY: USA
 : ZIP: 27709
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/998,416
 : FILING DATE: 24-DEC-1997
 : CLASSIFICATION: 435
 : PRIORITY APPLICATION DATA:
 : APPLICATION NUMBER: 01/0016/97
 : FILING DATE: 31-DEC-1996
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Mcigs, J. Timothy
 : REGISTRATION NUMBER: 38,241
 : REFERENCE/DOCKET NUMBER: PF/5-30306/A/CSC197b
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 919-541-8587

TELEFAX: 919-541-8684
 : INFORMATION FOR SEQ ID NO: 224:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 782 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : ORGANISM: FACI2080P
 : ORGANISM: FACI2080P
 : US-08-948-416-224

Query Match

Best Local Similarity 2.4%; Score 34.8; DB 4; Length 782;
 Matches 112; Conservative 0; Mismatches 107; Indels 0; Gaps 1;

QY 1247 taacacgttcttctgaatataatataatctctctgaacttctctgaacttctctgaact 1346
 Db 272 TAAATAGTATTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 341
 QY 1347 caatataatctctctgaacttctgaacttctgaacttctgaacttctgaacttctgaact 1409
 Db 342 TTAAGTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 388
 QY 1407 ctccac 1466
 Db 389 CTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 448
 QY 1467 ac 1506
 Db 448 AGTAATATTTTATTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 490

RESULT 3

US-08-475-035-4
 : Sequence 3, Application US/08475035
 : Patent No. 5985554
 : GENERAL INFORMATION:
 : APPLICANT: KING, C. R.
 : APPLICANT: KRAUS, MATTHIAS H.
 : APPLICANT: AMMONSON, STUART A.
 : TITLE OF INVENTION: HUMAN GENE RELATED TO HOT DISTINCT FLOW
 : TITLE OF INVENTION: EGF RECEPTOR GENE
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: NEEDLE & ROSENBERG, P.C.
 : STREET: Suite 1200, 127 Peachtree Street
 : CITY: Atlanta
 : STATE: Georgia
 : COUNTRY: USA
 : ZIP: 30304
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/475,035
 : FILING DATE: 7 Jun 1995
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Perryman, David G.
 : REGISTRATION NUMBER: 35,438
 : REFERENCE/DOCKET NUMBER: 1414, 159
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 404/688-0770
 : TELEFAX: 404/688-9880
 : INFORMATION FOR SEQ ID NO: 3:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 5542 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single

100

Copyright (c) 1990-1995 by Computer Aid.

OM nucleotide nucleic search using NW model

Run on: July 31, 2002 15:27:08 Search time 263.51 Seconds
(without alignments)
10263.02 Million cell updates/sec

Filter: US-09-844-502-1
Perfect score: 1575
Sequence: 1 tttaacaaagatctgcacatc ctttaacaaagatctgcacatc 1575

Scoring tables: IDENTITY_NW
Gapop 15.0 Gapext 3.0

Searches: 1736436 steps 4591722 successes
Total number of hits satisfying chosen parameters: 4472872

Minimum DB seq length: 6
Maximum DB seq length: 20000000

Post-processing: Minimum Match: 08
Maximum Match: 1004
Listing first 45 summaries

Database: NConos1032021*

Result No.	Score	Query Match	Length	Description
1	1575	100.0	1575	41 AAS5942
2	868.4	55.1	41	AAS5942
3	102	6.5	422	23 AAS6468
4	102	6.5	322	23 AAS6764
5	102	6.5	422	23 AAS6566
6	102	6.5	213	23 AAS9266
7	100.4	6.4	2104	23 AAS6468
8	100.4	6.4	1104	23 AAS6764
9	100.4	6.4	1104	23 AAS6566

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARYS

Result No.	Score	Query Match	Length	Description
1	1575	100.0	1575	41 AAS5942
2	868.4	55.1	41	AAS5942
3	102	6.5	422	23 AAS6468
4	102	6.5	322	23 AAS6764
5	102	6.5	422	23 AAS6566
6	102	6.5	213	23 AAS9266
7	100.4	6.4	2104	23 AAS6468
8	100.4	6.4	1104	23 AAS6764
9	100.4	6.4	1104	23 AAS6566

1	10	100.4	6.4	1104	23	AAS59389	DNA encoding novel
2	11	100.4	6.4	1305	23	AAS78571	DNA encoding novel
3	12	100.4	6.4	1305	23	AAS81062	DNA encoding novel
4	13	100.4	6.4	1553	23	AAS78576	DNA encoding novel
5	14	100.4	6.4	1553	23	AAS81066	DNA encoding novel
6	15	100.4	6.4	1564	23	AAS78572	DNA encoding novel
7	16	100.4	6.4	1564	23	AAS81063	DNA encoding novel
8	17	52	3.3	477	23	AAL01007	Human reproductive
9	18	39.6	2.5	2557	23	ABL10518	Drosophila melanog
10	19	38.8	2.5	1812	23	AAS78573	DNA encoding novel
11	20	38.8	2.5	1812	23	AAS81064	DNA encoding novel
12	21	38.8	2.5	2172	23	AAS6468	DNA encoding novel
13	22	38.8	2.5	2172	23	AAS6764	DNA encoding novel
14	23	38.8	2.5	2172	23	AAS85581	DNA encoding novel
15	24	38.8	2.5	2172	23	AAS9389	DNA encoding novel
16	25	38.8	2.5	2568	23	AAS79060	DNA encoding novel
17	26	38.6	2.5	4280	23	AAL4065	Yeast AD9604 asso
18	27	38.4	2.4	489	22	AHA4934	Human breast cell
19	28	38.4	2.4	489	22	AHA67852	Human fetal liver
20	29	38.4	2.4	489	22	AHA34911	Probe #13377 for g
21	30	38.4	2.4	489	22	AK16261	Human brain expres
22	31	38.4	2.4	489	22	AK42005	Human bone marrow
23	32	38.4	2.4	489	22	AK12720	Probe #12703 for q
24	33	38.4	2.4	489	22	AK148072	Probe #16758 used
25	34	38.4	2.4	489	22	AK108443	Probe #8434 used t
26	35	38.4	2.4	18193	22	AK87552	Human immune/haema
27	36	38.4	2.4	18252	22	AK87554	Human immune/haema
28	37	38.4	2.4	20188	22	AK73082	Human immune/haema
29	38	38.4	2.4	20188	22	AK87550	Human immune/haema
30	39	37.4	2.4	6126	22	AAS46573	Tumour suppressor
31	40	47.4	2.4	6126	24	ABL23832	Human immune syste
32	41	47.4	2.4	6531	41	AK32862	Human immune syste
33	42	47.4	2.4	34763	24	AAS46714	Tumour suppressor
34	43	47.2	2.4	1498	23	AAS6468	DNA encoding novel
35	44	47.2	2.4	1498	23	AAS6764	DNA encoding novel
36	45	47.2	2.4	1498	23	AAS78569	DNA encoding novel

ALIGNMENTS

RESULT 1
AAS59327 standard, cDNA; 1575 bp.
AAS59327:
97 REV 2000 (first entry)
cDNA encoding an oocyte maturation and proliferation protein (s-b,
oocyte maturation, cell division, cell proliferation, cancer,
cell differentiation, cell differentiation, cell differentiation,
cell cycle progression); ss.
Xenopus sp.
Key Location/Qualifiers
CDS 214..1116
/tag= a
/product= "IS26"

EP1026244-AL.
00-AUG-2000.
99-FEB-1999; 99EP-0102172.
99-FEB-1999; 99EP-0102172
(EMBL-) EMBL EURO LAB MOLEKULARBIOLOGIE.
WPI: 2000-516092/47.

XX	W0200175067-A2.
XX	11 (OCT-2001).
XX	30 MAR-2001; 2001WO-0508641.
XX	31-MAR-2000; 2000US-0540217.
XX	23-AUG-2000; 2000US-0649167.
XX	(HYSE-) HYSEQ INC.
XX	Dmanac RT, Liu C, Tang YT;
XX	WPI: 2001-639492/73.
XX	P-PSUB; ABG90505.
XX	New isolated polynucleotide and encoded polypeptides, useful in
XX	diagnostics, forensics, gene mapping, identification of mutations
XX	responsible for genetic disorders or other traits and to assess
XX	biodiversity
XX	Claim 1; SEQ ID No 496; 10pp; English.
XX	The invention relates to isolated polynucleotide (I) and
XX	polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX	and gene mapping, and in recombinant production of (II). The
XX	polynucleotides are also used in diagnostics as expressed sequence tags
XX	for identifying expressed genes. (I) is useful in gene therapy techniques
XX	to restore normal activity of (II) or to treat disease states involving
XX	(II). (II) is useful for generating antibodies against it, detecting or
XX	quantitating a polypeptide in tissue, as molecular weight markers and as
XX	a food supplement. (II) and its binding partners are useful in medical
XX	imaging of sites expressing (II). (I) and (II) are useful for treating
XX	disorders involving aberrant protein expression or biological activity.
XX	The polypeptide and polynucleotide sequences have applications in
XX	diagnostics, forensics, gene mapping, identification of mutations
XX	responsible for genetic disorders or other traits to assess biodiversity
XX	and to produce other types of data and products dependent on DNA and
XX	amino acid sequences. AAS64197-AAS94564 represent novel human
XX	diagnostic coding sequences of the invention
XX	Note: the sequence data for this patent did not appear in the printed
XX	specification, but was obtained in electronic format directly from WIPO
XX	at http://wipo.int/pub/published_pat_sequences .
XX	Sequence 1322 BP; 313 A; 453 C; 392 G; 264 T; 0 other;

```

Query Match      6.5%  Score 102; DB 23; Length 1322;
Best Local Similarity 66.3%  Pred. No. 5, 8e+21;
Matches 163; Conservative 0; Mismatches 80; Indels 3; Gaps 1;

QY 389 tacagccacagagccagcagcctctacaggctctctagaaaatqacagatcaggaaat 448
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 640 *gctccctgagacccagagcctctacaggctctctgaggtctctctctctaaagat 699
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 449 tctttctatgaactccctgctaaagattctccacaaatctccctaaatgaatctctaa 508
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 700 tcttgacctggacaaagatctcagggtctgagacagatctccctggataggctaaag 759
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 509 catattttaagggcagcagcctctacaccagcaggtacacaaatgaattctcttq 568
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 760 catattttaagcagcagcctctccctccctccctccctccctccctccctccctcc 816
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 569 tgcctctatcttgcctaatgacatgagagaatgaagaagatataaataatgaatct 628
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 tggctctctactggcctaataatgacatgagagagacacagaggtctccaaacaaacatct 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 629 tccctct 634
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 877 tccact 882
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

XX	AA0567034 standard; cDNA; 1322 bp.
XX	AA567034;
XX	13 FEB 2002 (first entry)
XX	DNA encoding novel human diagnostic protein #2808.
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorders.
XX	Homo sapiens.
XX	W020017067-A2.
XX	11 OCT-2001.
XX	30 MAR-2001; 2001WO 05066 A1
XX	31 MAR-2000; 2000US 954213.
PR	23 AUG 2000; 2000US 9649167.
XX	(HYTE) HYSEQ INC.
PA	Branan RI, Liu C, Tama YT;
PI	WO/01/63946-273.
IN	P FSOB; ABG02847.
XX	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PI	biodiversity
XX	Claim 1: SEQ ID No 2838; 104pp; English.
XX	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridization probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosomal
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in strategies as expected sequence to
CC	for identifying expressed genes. (I) is useful in gene therapy tech-
CC	niques to restore normal activity of (II) or to treat disease states involv-
CC	ing (II). (II) is useful for generating antibodies against it, detecting and
CC	quantitating a polypeptide in tissue, as molecular weight markers and
CC	a food supplement. (II) and its binding partners are useful in med-
CC	icating of sites expression (II). (I) and (II) are useful in med-
CC	ical disorders involving aberrant protein expression or biological activ-
CC	ity. The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiver-
CC	sity and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AA564197 AA564564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	Specification, but was obtained in electronic format directly from WIPO
CC	at http://wipo.int/pub/published.pat.sequences .
XX	Sequence 1322 BP; 31 A; 33 C; 92 G; 264 T; 0 other.

[illegible]




```

AJ249978.1  GI:606823
ls26 gene; p31 rino.
SOURCE      African clawed frog.
ORGANISM    Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 1376)
Feiby, I., Blazquez, M., Palmer, A., Ertija, R. and Nebreda, A. R.
A novel p31/cdc2-binding and activating protein that is necessary
and sufficient to triage G(2)/M progression in Xenopus oocytes
Genes Dev. 13 (16), 2177-2189 (1999)
99396721
2 (bases 1 to 1376)
Nebreda, A. R.
Direct Submission
Submitted (12-OCT-1999) Nebreda, A. R. | Developmental Biology
Programme, EMBL, Meyerhofstrasse 1, 69117 Heidelberg, GERMANY
location/Qualifiers
1. 1376
/oranism="Xenopus laevis"
/db_xref="taxon:8355"
214. 1113
/gene="ls26/clone 5"
214. 1113
/gene="ls26/clone 5"
/function="triages oocyte meiotic maturation"
/codon_start=1
/product="p31 rino"
/protein_id="CA58366.1"
/db_xref="GI:606824"
/translation="MKHMSVIRATSGSGVKVVGCKGHVHARVVCARKAQLPPEE
LSVKPQWENTILNLPQERQAFYELLFNEQLQELSMQSLPISPKYLINMVLATPE
RAGLYSEITTNFVALYLANDMEEDDEKYELFPWALGDSWRELFPGFQLRDLDF
WAKMTTRAVSRGCRTEVSKSPKTHAWLPEPTTUESAMPQYIPNEDCTSTGGELT
PASCITLCPAGVSGYSGVSHNSNSDQETPHYINRQWSQELLMSPELLLDPECHD
LRLQPEVGLGFEDCTALFWHH."
358 a 331 c 336 g 351 t
BASE COUNT
ORIGIN

```

Query Match: 82.0%; Score 1291.2; DB 5; Length 1376;	
Host Local Similarity 98.7%; Prod. No. 0;	
Matches 1346; Conservative 0; Mismatches 8	Indels 10; Gaps 4;
1	tttaacaaqqacttgcagctccagatgtatgttttttttcadaagctccgctccaaatgctatatt 60
1	TTAAACAGGACTTCCAGCTCCAGCTGAGCTTTTTCACAAGCTCCGCTTAAAGCTGCTAT 60
61	ttttttattttccagaaagctatataagaaagacagacataaagaaagttgagcagaattctctg 120
61	TTTTTTTATTATCCAGAGAGCTATATAAAGAGCAAGCAAGAGAAATAGAGGAGATTCCTG 120
121	ttttatgcgaatttcggcccaatgattggttaagcgcctgaaggttgtatgtcttttgtctcag 180
121	TTTTATGCCATTTCGGCAGTGGTGGCTAAGGCTCTCAAGCTGGCTGTCTCTTGGCTCAG 180
181	atcaacccctcgagcccaattgtctccctttctacaaataagagacattgcgaattataacccag 240
181	ATCAACCCCTGGGCCCGCTGTGCCCTTCTACAAATAGAGCACAATGACAGTGTAAACCCG 240
241	gcagagctccattttgtggcagcgaggtgaacgaagttcatttggcgaatgataatccgacaccc 300
241	GCAACCTCCATTTTGTGGCAGCGGGGTGAAGCAGAGTCAATTTGCAAGAGCTATGCAAGCT 300
301	cggagttgtttgagcgcgcgcgaagcgcaaatccctgaagagagaggaagtttgtcagttcdaaccc 360
301	CGGCTTTGTTGGACCGCGCAAGCGCAATTCCTTGAGACAGAGTACGTTCTAGTCAAACTG 360
361	aaatagtagcgaataatcccaattctcaattctacagaccccaagaaagcagagctctctacacag 420
361	AAATAGTTCGAAATATCTATCTCAATCTACAGCCGAGGAGCTGTACGCTCTTCTACAG 420

RESULT 4
XLA133500
LACUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE

XLA134500
Xenopus laevis mRNA
A134500
1527 bp; cDNA clone
ATCC14619

1 kb; bp
p34.1 (1.2)

mRNA
VPL 14.14.200

[illegible]

ID 67442 TCTTCTTTGTTAGTATCTGTTATATACAGTTGAGAGAGALGAAAGAAAGAGAGAG 67383
 QY 623 aaatcttccctgggaactaggdaactcgtggctgagcttttccgacaatttttgcac 682
 DB 67382 AAAATTTTCCATGGCTTTAGGCAAAAACAGCAGAAAATCTCTGCTAAATCTTAAACT 67323
 QY 683 tcccaacacattcttcaatcaaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 742
 DB 67322 TAAGGACACAGCTCGGATACAAATCTACATATAGGCTATCTTAAAGACGAGCTGTG 67263
 QY 743 atgaatgaat 751
 DB 67262 AGGAGATTA 67254
 RESULT 10
 LOCUS AC105890 190672 bp DNA linear HTG 10-JAN-2002
 DEFINITION Rattus norvegicus clone CH230 159L8, *** sequence in progress
 *** 110 unordered pieces.
 ACCESSION AC105890
 VERSION AC105890.1 GI:18104797
 KEYWORDS HTG; HTGS_PHASE1
 SOURCE Norway rat
 ORGANISM Rattus norvegicus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 190672)
 Muzny, D.M., Adams, C., Adio-oduola, B., Ali-osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amaralunpe, B.C., Are, J.R., Banks, T., Barbara, J.,
 Benton, J., Bimaga, K., Blankenbuck, K., Bonnin, D., Bourk, J.,
 Bowie, S., Brice, M., Brown, R., Bryant, N.P., Bryant, N.P., Huhay, C.,
 Hurch, P., Burkett, C., Hurrell, K.L., Hyrd, N.C., Carron, T.F.,
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
 Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
 Coyle, M.D., Dalhorne, S.R., David, R., Davila, M.L., Davis, C.,
 Davy, Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
 Denn, A.L., Dinh, Y., Dinh, H.H., Douthett, K.J., Draper, H.,
 Duhan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
 Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A.,
 Hernandez, J., Hernandez, O., Hodgeson, A., Hoque, M., Hollway, C.,
 Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
 Jackson, L.R., Jacobson, R., Jia, Y., Johnson, K., Jolivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, F., Kind, L., Korvach, J.,
 Kovar, C., Kraus, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, J., Liu, J., Liu, W.,
 Louisse, H., Lozano, K.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, K., Mayhew, A.,
 Martinez, F., Massey, F., Mawhinney, E., Melrod, M.P., Meador, M.,
 Mei, G., Melzer, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morahan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
 Ogih, M., Okunishi, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Percy, J., Perez, I., Peters, J., Pickens, K., Pimus, E., Pohl, J.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojehkan, I., Rolfe, M.,
 Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shostakov, N.,
 Slismon, I., Sutton, E., Souda, T., Sparks, A., Stanley, J.,
 Stone, H., Tansoy, J., Taylor, C., Tabor, P., Tamerisa, A., Tamerisa, K.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 190672)
 REFERENCE
 Worley, K.C.
 Direct Submission
 TITLE
 JOURNAL
 Unpublished
 REFERENCE
 Worley, K.C.
 Direct Submission
 TITLE

JOURNAL

COMMENT

Submitted (10 JAN 2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: CNZO
 Center clone name: CH230 159L8
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329; first call to
 findPhrapList
 Consensus quality: 140479 bases at least Q40
 Consensus quality: 153171 bases at least Q40
 Consensus quality: 166347 bases at least Q20
 Estimated insert size: 12027; sum of contigs estimation
 quality coverage: 6x in Q20 bases; average bp estimation
 quality coverage: 1.1x in Q20 bases; sum of contigs estimation

* Note: Estimated insert size may differ from sequence length
 * (see <http://www.bcm.tmc.edu/hgsc/Phrap/Phrap.html>)
 * NOTE: This is a working draft sequence. It currently
 * consists of 110 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 3698: contig of 4698 bp in length
 3798: gap of unknown length
 3799: contig of 4577 bp in length
 8376: gap of unknown length
 8476: contig of 2747 bp in length
 11212: gap of unknown length
 11213: contig of 4401 bp in length
 15614: gap of unknown length
 15714: contig of 3143 bp in length
 18957: gap of unknown length
 22789: contig of 8833 bp in length
 22890: gap of unknown length
 24375: contig of 1486 bp in length
 24476: gap of unknown length
 26148: contig of 1663 bp in length
 26149: gap of unknown length
 29106: contig of 2808 bp in length
 29206: gap of unknown length
 31666: contig of 2460 bp in length
 31767: gap of unknown length
 33524: contig of 1758 bp in length
 34625: gap of unknown length
 36095: contig of 2471 bp in length
 36096: gap of unknown length
 36195: contig of 2635 bp in length
 38830: gap of unknown length
 38930: contig of 2297 bp in length
 41227: gap of unknown length
 41328: contig of 2458 bp in length
 43288: gap of unknown length
 43786: contig of 2902 bp in length
 44886: gap of unknown length
 45987: contig of 2559 bp in length
 45988: gap of unknown length
 45989: contig of 2949 bp in length
 48647: gap of unknown length
 51595: gap of unknown length
 51596: contig of 1750 bp in length
 53645: gap of unknown length
 53745: contig of 1055 bp in length
 54800: gap of unknown length
 54900: contig of 1478 bp in length
 54901: gap of unknown length
 56478: gap of unknown length

RESULT 12

AP004068/c

LOCUS

DEFINITION

Homo sapiens chromosome 11 clone RP11-399J13 map 11q, WORKING DRAFT

SEQUENCE, 48 unordered pieces

AP004068

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AP004068 170658 bp PNA 11q:0.07-0.08 2002
 Homo sapiens chromosome 11 clone RP11-399J13 map 11q, WORKING DRAFT
 SEQUENCE, 48 unordered pieces
 AP004068
 HTG: HTGS, PHASE1, HTGS_DRAFT
 Homo sapiens DNA, clone:RP11-399J13.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hongo-Seoq,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Homo sapiens 170,658 genomic DNA of 11q
 Published only in database (2001) in press
 2 (bases 1 to 170658)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hongo-Seoq,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Direct Submission
 Submitted (28-DEC-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:hattori@gsc.riken.go.jp, rpl:htp://hqp.gsc.riken.go.jp/
 Tel:81-45-503-9111 Fax:81-45-503-9170)
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: http://hqp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft11
 Center clone name: RP11-399J13
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990429
 Consensus quality: 157557 bases at least Q40
 Consensus quality: 162348 bases at least Q30
 Consensus quality: 164588 bases at least Q20
 Insert size: 165958; sum-of-contigs
 Quality coverage: 8.38x in Q20 bases; sum-of-contigs

NOTE:

This is a 'working draft' sequence. It currently consists of 48 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

1 10293: contig of 10293 bp in length
 10294 10493: gap of 100 bp
 10394 21401: contig of 11008 bp in length
 21402 21501: gap of 100 bp
 21502 29976: contig of 8475 bp in length
 29977 30076: gap of 100 bp
 30077 38923: contig of 8847 bp in length
 38924 39023: gap of 100 bp
 39024 43854: contig of 4831 bp in length
 43855 43954: gap of 100 bp
 43955 51879: contig of 7925 bp in length
 51880 51979: gap of 100 bp
 51980 57476: contig of 5397 bp in length
 57477 63579: contig of 6103 bp in length
 63580 63679: gap of 100 bp
 63680 68235: contig of 4556 bp in length
 68236 68335: gap of 100 bp
 68336 73407: contig of 5072 bp in length
 73408 73507: gap of 100 bp
 73508 78481: contig of 4974 bp in length
 78482 78581: gap of 100 bp
 78582 84400: contig of 4819 bp in length
 84401 83500: gap of 100 bp
 84501 88466: contig of 4966 bp in length
 88467 88566: gap of 100 bp
 88567 91704: contig of 4148 bp in length
 91705 91804: gap of 100 bp
 91805 96027: contig of 4223 bp in length
 96028 96127: gap of 100 bp
 96128 99548: contig of 4421 bp in length
 99549 99648: gap of 100 bp
 99649 103711: contig of 4063 bp in length
 103712 103811: gap of 100 bp
 103812 108909: contig of 5098 bp in length
 108910 109009: gap of 100 bp
 109010 111514: contig of 2505 bp in length
 111515

